(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 January 2002 (24.01.2002)

PCT

(10) International Publication Number
WO 02/06446 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US01/22523

(22) International Filing Date: 17 July 2001 (17.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/218,727 17 July 2000 (17.07.2000) US

(71) Applicant *(for all designated States except US)*: **GENAISSANCE PHARMACEUTICALS, INC.** [US/US]; A Corporation of the State of Delaware, Five Science Park, New Haven, CT 06511 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **KLIEM, Stefanie, E.** [DE/DE]; Kiefernweg 37, 61440 Oberusel (DE). **KOSHY, Beena** [IN/US]; 1298 Hartford Turnpike, Apartment 11B, North Haven, CT 06473 (US).

(74) Agents: **FIELD, Gisela, M.** et al.; Genaissance Pharmaceuticals, Inc., Five Science Park, New Haven, CT 06511 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/06446 A2

(54) Title: HAPLOTYPES OF THE EDG6 GENE

(57) Abstract: Novel genetic variants of the Endothelial Differentiation, G-Protein-Coupled Receptor 6 (EDG6) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the EDG6 gene. Compositions and methods for haplotyping and/or genotyping the EDG6 gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.

HAPLOTYPES OF THE EDG6 GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/218,727 Serial No. filed July 17, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human endothelial differentiation, G-protein-coupled receptor 6 (EDG6) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of such drugs in clinical trials or their early withdrawal from the market even though they could be highly beneficial for

other groups in the population. This problem significantly increases the time and cost of drug discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* **15**:1249-52; Kleyn PW et al. 1998 *Science* **281**: 1820-21; Kola I 1999 *Curr Opin Biotech* **10**:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* **66**:445-7; Marshall, E 1999 *Science* **284**:406-7; Judson R et al. 2000 *Pharmacogenomics* **1**:1-12; Roses AD 2000 *Nature* **405**:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* **19**:216-7; Wang DG et al 1998 *Science* **280**:1077-82; Chakravarti A 1999 *Nat Genet* **21**:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* **4**:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* **5**:538-43; Davidson S 2000 *Nature Biotech* **18**:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* **161**: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* **320**:987-90; Dahl BS 1997 *Acta Psychiatr Scand* **96** (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* **63**:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* **97**:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of cancer, angiogenesis and inflammation is the endothelial differentiation, G-protein-coupled receptor 6 (EDG6) gene or its encoded product. EDG receptors, such as EDG6, constitute a novel subfamily of G-protein-coupled receptors displaying a heterogeneous expression pattern. Members of this family can bind lysophospholipids or lysosphingolipids as ligands. EDG6 is specifically expressed in fetal and adult lymphoid and hematopoietic tissue as well as in lung (Graler et al., *Genomics* 1998; 53:164-169). Graler et al. (*supra*) suggest that because of the known mitogenic and chemotactic activity of bioactive

lipids, EDG6 may play an essential role in lymphocyte cell signaling. EDG6 can also bind sphingosine 1-phosphate, a lysolipid, to elicit biological responses, including mitogenesis, differentiation, migration and apoptosis, via receptor-dependent mechanisms. Sphingosine 1-phosphate has been implicated in pathophysiological disease states, such as cancer, angiogenesis and inflammation (Pyne and Pyne, *Biochem J* 2000; 349:385-402). For example, sphingosine 1-phsophate (S1-P) has been shown to induce the secretion of type H Insulin-like growth factor II, which is responsible for proliferation of cultured breast cancer cells (Goetzl et al., *Cancer Res.* 1999; 59:4732-4737). Goetzl et al. have shown that another EDG receptor, EDG4, is a marker for ovarian cancer, and it is possible that other S1-P-specific EDG receptors may be involved in cancer. Therefore, aberrant expression of EDG6 may result in changes in S1-P concentrations, which could affect several disease processes.

The endothelial differentiation, G-protein-coupled receptor 6 gene is located on chromosome 19p13.3 and contains 1 exon that encodes a 384 amino acid protein. A reference sequence for the EDG6 gene is shown in the contiguous lines of Figure 1 (Genaissance Reference No. 3216828; SEQ ID NO: 1). Reference sequences for the coding sequence (GenBank Accession No. NM_003775.1) and protein are shown in Figures 2 (SEQ ID NO: 2) and 3 (SEQ ID NO: 3), respectively.

Because of the potential for variation in the EDG6 gene to affect the expression and function of the encoded protein, it would be useful to know whether polymorphisms exist in the EDG6 gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of EDG6 as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 23 novel polymorphic sites in the EDG6 gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 3591 (PS1), 3697 (PS2), 3804 (PS3), 3818 (PS4), 4123 (PS5), 4240 (PS6), 4472 (PS7), 4499 (PS8), 4531 (PS9), 4574 (PS10), 4736 (PS11), 4813 (PS12), 5068 (PS13), 5103 (PS14), 5150 (PS15), 5179 (PS16), 5301 (PS17), 5333 (PS18), 5448 (PS19), 5560 (PS20), 5580 (PS21), 5587 (PS22) and 5606 (PS23). The polymorphisms at these sites are guanine or adenine at PS1, cytosine or thymine at PS2, cytosine or thymine at PS3, adenine or guanine at PS4, cytosine or thymine at PS5, guanine or adenine at PS6, guanine or adenine at PS7, guanine or adenine at PS8, guanine or adenine at PS9, guanine or thymine at PS10, cytosine or thymine at PS11, cytosine or thymine at PS12, cytosine or thymine at PS13, guanine or thymine at PS14, guanine or adenine at PS15, guanine or adenine at PS16, guanine or adenine at PS17, guanine or adenine at PS18, guanine or cytosine at PS19, guanine or adenine at PS20, guanine or adenine at PS21, cytosine or thymine at PS22 and guanine or cytosine at PS23. In addition, the inventors have determined the identity of the alleles at these sites in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of

haplotypes and haplotype pairs for PS1-PS23 in the EDG6 gene, which are shown below in Tables 5 and 4, respectively. Each of these EDG6 haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the EDG6 gene that exists in the human population. The frequency with which each haplotype and haplotype pair occurs within the total reference population and within each of the four major population groups included in the reference population was also determined.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the EDG6 gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22 and PS23 in both copies of the EDG6 gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel EDG6 polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel EDG6 polymorphic sites. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 5 below or has one of the haplotype pairs in Table 4 below.

The invention also provides a method for haplotyping the EDG6 gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the EDG6 gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22 and PS23. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's EDG6 gene is defined by one of the EDG6 haplotypes shown in Table 5, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's EDG6 gene are defined by one of the EDG6 haplotype pairs shown in Table 4 below, or a sub-haplotype pair thereof. The method for establishing the EDG6 haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with EDG6 activity, e.g., cancer, angiogenesis and inflammation.

For example, the haplotyping method can be used by the pharmaceutical research scientist to validate EDG6 as a candidate target for treating a specific condition or disease predicted to be associated with EDG6 activity. Determining for a particular population the frequency of one or more of the individual EDG6 haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue EDG6 as a target for treating the specific disease of interest. In particular, if variable EDG6 activity is associated with the disease, then one or more EDG6 haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed EDG6 haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable EDG6 activity has little, if any,

involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any EDG6 haplotype or haplotype pair, apply the information derived from detecting EDG6 haplotypes in an individual to decide whether modulating EDG6 activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting EDG6 to treat a specific condition or disease predicted to be associated with EDG6 activity. For example, detecting which of the EDG6 haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the most frequent EDG6 isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular EDG6 haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

The method for haplotyping the EDG6 gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with EDG6 activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the EDG6 haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute EDG6 haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of an EDG6 haplotype or haplotype pair that had a previously unknown association with response to the drug being studied in the trial. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any EDG6 haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association between a trait and an EDG6 genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the EDG6 genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the EDG6 genotype or haplotype in a reference population. A higher frequency of the EDG6 genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the EDG6 genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the EDG6 haplotype is selected from the haplotypes shown in Table 5, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for cancer, angiogenesis and inflammation.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the EDG6 gene or a fragment thereof. The reference sequence comprises the contiguous sequences shown in Figure 1 and

the polymorphic variant comprises at least one polymorphism selected from the group consisting of adenine at PS1, thymine at PS2, thymine at PS3, guanine at PS4, thymine at PS5, adenine at PS6, adenine at PS7, adenine at PS8, adenine at PS9, thymine at PS10, thymine at PS11, thymine at PS12, thymine at PS13, thymine at PS14, adenine at PS15, adenine at PS16, adenine at PS17, adenine at PS18, cytosine at PS19, adenine at PS20, adenine at PS21, thymine at PS22 and cytosine at PS23.

A particularly preferred polymorphic variant is an isogene of the EDG6 gene. An EDG6 isogene of the invention comprises guanine or adenine at PS1, cytosine or thymine at PS2, cytosine or thymine at PS3, adenine or guanine at PS4, cytosine or thymine at PS5, guanine or adenine at PS6, guanine or adenine at PS7, guanine or adenine at PS8, guanine or adenine at PS9, guanine or thymine at PS10, cytosine or thymine at PS11, cytosine or thymine at PS12, cytosine or thymine at PS13, guanine or thymine at PS14, guanine or adenine at PS15, guanine or adenine at PS16, guanine or adenine at PS17, guanine or adenine at PS18, guanine or cytosine at PS19, guanine or adenine at PS20, guanine or adenine at PS21, cytosine or thymine at PS22 and guanine or cytosine at PS23. The invention also provides a collection of EDG6 isogenes, referred to herein as an EDG6 genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for an EDG6 cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of thymine at a position corresponding to nucleotide 114, adenine at a position corresponding to nucleotide 231, adenine at a position corresponding to nucleotide 463, adenine at a position corresponding to nucleotide 490, adenine at a position corresponding to nucleotide 522, thymine at a position corresponding to nucleotide 565, thymine at a position corresponding to nucleotide 727, thymine at a position corresponding to nucleotide 804, thymine at a position corresponding to nucleotide 1059, thymine at a position corresponding to nucleotide 1094 and adenine at a position corresponding to nucleotide 1141. A particularly preferred polymorphic cDNA variant comprises the coding sequence of an EDG6 isogene defined by haplotypes 3c, 7c-12c, 19c-22c, and 24c.

Polynucleotides complementary to these EDG6 genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the EDG6 gene will be useful in studying the expression and function of EDG6, and in expressing EDG6 protein for use in screening for candidate drugs to treat diseases related to EDG6 activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express EDG6 for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the EDG6 protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises at least one variant amino acid

selected from the group consisting of arginine at a position corresponding to amino acid position 155, serine at a position corresponding to amino acid position 164, serine at a position corresponding to amino acid position 189, cysteine at a position corresponding to amino acid position 243, leucine at a position corresponding to amino acid position 365 and methionine at a position corresponding to amino acid position 381. A polymorphic variant of EDG6 is useful in studying the effect of the variation on the biological activity of EDG6 as well as on the binding affinity of candidate drugs targeting EDG6 for the treatment of cancer, angiogenesis and inflammation.

The present invention also provides antibodies that recognize and bind to the above polymorphic EDG6 protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one of the EDG6 polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the EDG6 isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against EDG6 protein, and for testing the efficacy of therapeutic agents and compounds for cancer, angiogenesis and inflammation in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the EDG6 gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the EDG6 gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing EDG6 haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the EDG6 gene (Genaissance Reference No. 3216828; contiguous lines), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:1 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25). SEQ ID NO:119 is a modified version of SEQ ID NO:1 that shows the context sequence of each polymorphic site, PS1-PS23, in a uniform format to facilitate electronic searching. For each polymorphic site, SEQ ID NO:119 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each PS is separated by genomic sequence whose composition is defined elsewhere herein.

Figure 2 illustrates a reference sequence for the EDG6 coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a

reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the EDG6 protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the EDG6 gene. As described in more detail below, the inventors herein discovered 24 isogenes of the EDG6 gene by characterizing the EDG6 gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (21 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference population were organized into population subgroups by their self-identified ethnogeographic origin as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		21
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		18
	Caribbean	8
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the

CEPH Utah cohort) and one two-generation African-American family.

The EDG6 isogenes present in the human reference population are defined by haplotypes for 23 polymorphic sites in the EDG6 gene, all of which are believed to be novel. The novel EDG6 polymorphic sites identified by the inventors are referred to as PS1-PS23 to designate the order in which they are located in the gene (see Table 3 below). Using the genotypes identified in the Index Repository for PS1-PS23 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the EDG6 gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the EDG6 gene include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether EDG6 is a suitable target for drugs to treat cancer, angiogenesis and inflammation, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene – A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype – An unphased 5′ to 3′ sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype – The unphased 5′ to 3′ sequence of nucleotide pairs found at all polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype – The unphased 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Genotyping – A process for determining a genotype of an individual.

Haplotype – A 5′ to 3′ sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype – The 5′ to 3′ sequence of nucleotides found at all polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Sub-haplotype – The 5′ to 3′ sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Haplotype pair – The two haplotypes found for a locus in a single individual.

Haplotyping – A process for determining one or more haplotypes in an individual and includes

use of family pedigrees, molecular techniques and/or statistical inference.

Haplotyp data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment - A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the EDG6 gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel EDG6 polymorphisms and haplotypes identified herein.

The compositions comprise at least one EDG6 genotyping oligonucleotide. In one embodiment, an EDG6 genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art,

including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of an EDG6 polynucleotide, i.e., an EDG6 isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-EDG6 polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the EDG6 gene using the polymorphism information provided herein in conjunction with the known sequence information for the EDG6 gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5′ end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th position in a 20mer). An ASO primer of the invention has a 3′ terminal nucleotide, or preferably a 3′ penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention.

ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting EDG6 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

GTGTGCTRAGCGCCG (SEQ ID NO:4) and its complement,
GGCCCATYCCGAGTG (SEQ ID NO:5) and its complement,
GGGGGTCYTCACAGC (SEQ ID NO:6) and its complement,
CCAGGGCRGCCCCAG (SEQ ID NO:7) and its complement,
CCGGGCGYGGGGGGC (SEQ ID NO:8) and its complement,
TGCGGTCRCGACGCT (SEQ ID NO:9) and its complement,
CGAGAGCRGGGCCAC (SEQ ID NO:10) and its complement,
CGTCTACRGCTTCAT (SEQ ID NO:11) and its complement,
TGGCCGCRCTGCTGG (SEQ ID NO:12) and its complement,
CCTGTGCKCCTTTGA (SEQ ID NO:13) and its complement,
AGCGGCCYGCCGCAA (SEQ ID NO:14) and its complement,
CACTCTTYGGGCTGC (SEQ ID NO:15) and its complement,
CCGACAGYTCTCTGA (SEQ ID NO:16) and its complement,
GGCTCCCKCTCGCTC (SEQ ID NO:17) and its complement,
CTCCAGCRTGCGGAG (SEQ ID NO:18) and its complement,
GTCTTGCRTGTGGAT (SEQ ID NO:19) and its complement,
TCTTCCCRGTGGCCT (SEQ ID NO:20) and its complement,
CAAATGGRCTTCCCA (SEQ ID NO:21) and its complement,
GATTCTGSGGAAGTC (SEQ ID NO:22) and its complement,
ATGTTGCRGCCTCTT (SEQ ID NO:23) and its complement,
CTGGTGCRTGCATGC (SEQ ID NO:24) and its complement,
GTGCATGYGTGGGGG (SEQ ID NO:25) and its complement, and
GGCTCAGSGGGGCTG (SEQ ID NO:26) and its complement.

A preferred ASO primer for detecting EDG6 gene polymorphisms comprises a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

CCTGCTGTGTGCTRA (SEQ ID NO:27); CTCCACCGGCGCTYA (SEQ ID NO:28);
AGGGGTGGCCCATYC (SEQ ID NO:29); AGTCCCCACTCGGRA (SEQ ID NO:30);
AGGGGTGGGGGTCYT (SEQ ID NO:31); GCCCTGGCTGTGARG (SEQ ID NO:32);
TCACAGCCAGGGCRG (SEQ ID NO:33); AACGCGCTGGGGCYG (SEQ ID NO:34);
GGCTGGCCGGGCGYG (SEQ ID NO:35); CCTCCGGCCCCCCRC (SEQ ID NO:36);
GCCACATGCGGTCRC (SEQ ID NO:37); AGACCCAGCGTCGYG (SEQ ID NO:38);
GGTGGCCGAGAGCRG (SEQ ID NO:39); GTCTTGGTGGCCCYG (SEQ ID NO:40);
CAGCCGCGTCTACRG (SEQ ID NO:41); AGGCCGATGAAGCYG (SEQ ID NO:42);

GGCTGCTGGCCGCRC (SEQ ID NO:43); GCATCCCCAGCAGYG (SEQ ID NO:44);
GAACTGCCTGTGCKC (SEQ ID NO:45); CAGCGGTCAAAGGMG (SEQ ID NO:46);
ACGCCCAGCGGCCYG (SEQ ID NO:47); CGGGCCTTGCGGCRG (SEQ ID NO:48);
GGGGCCCACTCTTYG (SEQ ID NO:49); CCAGCAGCAGCCCRA (SEQ ID NO:50);
CCACCACCGACAGYT (SEQ ID NO:51); TTGGCCTCAGAGARC (SEQ ID NO:52);
TTTCGCGGCTCCCKC (SEQ ID NO:53); AAAGCTGAGCGAGMG (SEQ ID NO:54);
CAGCATCTCCAGCRT (SEQ ID NO:55); CAGATGCTCCGCAYG (SEQ ID NO:56);
GTTGCAGTCTTGCRT (SEQ ID NO:57); TGCACCATCCACAYG (SEQ ID NO:58);
CCATGGTCTTCCCRG (SEQ ID NO:59); CCCGAGAGGCCACYG (SEQ ID NO:60);
TGACGCCAAATGGRC (SEQ ID NO:61); TGACCATGGGAAGYC (SEQ ID NO:62);
CTGTGTGATTCTGSG (SEQ ID NO:63); GGCCGGGACTTCCSC (SEQ ID NO:64);
TACGTGATGTTGCRG (SEQ ID NO:65); GGGAATAAGAGGCYG (SEQ ID NO:66);
TATTCCCTGGTGCRT (SEQ ID NO:67); CCCCACGCATGCAYG (SEQ ID NO:68);
TGGTGCGTGCATGYG (SEQ ID NO:69); CCACGGCCCCCACRC (SEQ ID NO:70);
GGCCGTGGCTCAGSG (SEQ ID NO:71); and GATCCACAGCCCCSC (SEQ ID NO:72).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3′-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting EDG6 gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5′ to 3′, selected from the group consisting of:

GCTGTGTGCT (SEQ ID NO:73); CACCGGCGCT (SEQ ID NO:74);
GGTGGCCCAT (SEQ ID NO:75); CCCCACTCGG (SEQ ID NO:76);
GGTGGGGGTC (SEQ ID NO:77); CTGGCTGTGA (SEQ ID NO:78);
CAGCCAGGGC (SEQ ID NO:79); GCGCTGGGGC (SEQ ID NO:80);
TGGCCGGGCG (SEQ ID NO:81); CCGGCCCCCC (SEQ ID NO:82);
ACATGCGGTC (SEQ ID NO:83); CCCAGCGTCG (SEQ ID NO:84);
GGCCGAGAGC (SEQ ID NO:85); TTGGTGGCCC (SEQ ID NO:86);
CCGCGTCTAC (SEQ ID NO:87); CCGATGAAGC (SEQ ID NO:88);
TGCTGGCCGC (SEQ ID NO:89); TCCCCAGCAG (SEQ ID NO:90);
CTGCCTGTGC (SEQ ID NO:91); CGGTCAAAGG (SEQ ID NO:92);
CCCAGCGGCC (SEQ ID NO:93); GCCTTGCGGC (SEQ ID NO:94);
GCCCACTCTT (SEQ ID NO:95); GCAGCAGCCC (SEQ ID NO:96);
CCACCGACAG (SEQ ID NO:97); GCCTCAGAGA (SEQ ID NO:98);
CGCGGCTCCC (SEQ ID NO:99); GCTGAGCGAG (SEQ ID NO:100);
CATCTCCAGC (SEQ ID NO:101); ATGCTCCGCA (SEQ ID NO:102);
GCAGTCTTGC (SEQ ID NO:103); ACCATCCACA (SEQ ID NO:104);
TGGTCTTCCC (SEQ ID NO:105); GAGAGGCCAC (SEQ ID NO:106);
CGCCAAATGG (SEQ ID NO:107); CCATGGGAAG (SEQ ID NO:108);
TGTGATTCTG (SEQ ID NO:109); CGGGACTTCC (SEQ ID NO:110);
GTGATGTTGC (SEQ ID NO:111); AATAAGAGGC (SEQ ID NO:112);
TCCCTGGTGC (SEQ ID NO:113); CACGCATGCA (SEQ ID NO:114);
TGCGTGCATG (SEQ ID NO:115); CGGCCCCCAC (SEQ ID NO:116);
CGTGGCTCAG (SEQ ID NO:117); and CCACAGCCCC (SEQ ID NO:118).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

EDG6 genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized EDG6 genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the EDG6 gene in an individual. As used herein, the terms "EDG6 genotype" and "EDG6 haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the EDG6 gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid sample comprising the two copies of the EDG6 gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22 and PS23 in the two copies to assign an EDG6 genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-PS23.

Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may

be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the EDG6 gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5′ and 3′ untranslated regions. If an EDG6 gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the EDG6 gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22 and PS23 in that copy to assign an EDG6 haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the EDG6 gene or fragment such as one of the methods described above for preparing EDG6 isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two EDG6 gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional EDG6 clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the EDG6 gene in an individual. In a particularly preferred embodiment, the nucleotide at each of PS1-PS23 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the EDG6 haplotypes shown in Table 5. This can be accomplished by identifying, for one or both copies of the individual's EDG6 gene, the phased sequence of nucleotides present at each of PS1-PS23. The present invention also contemplates that typically only a subset of PS1-PS23 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 5. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* **97**:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* **22**:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* **4**:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, an EDG6 haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22 and PS23 in each copy of the EDG6 gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS23 in each copy of the EDG6 gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy

of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the EDG6 gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* **89**:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the EDG6 gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA* 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping another polymorphic site that is in linkage

disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

In another aspect of the invention, an individual's EDG6 haplotype pair is predicted from its EDG6 genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying an EDG6 genotype for the individual at two or more EDG6 polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing EDG6 haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the EDG6 haplotype pairs shown in Table 4.

Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a q% chance of not missing a haplotype that exists in the population at a p% frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n=\log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size

does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting an EDG6 haplotype pair for an individual, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*). A preferred process for predicting EDG6 haplotype pairs from EDG6 genotypes is described in U.S. Provisional Application Serial No. 60/198,340 and the corresponding International Application, PCT/US01/12831.

The invention also provides a method for determining the frequency of an EDG6 genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel EDG6 polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be a reference population, a family population, a same sex population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for EDG6 genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and an EDG6 genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive

genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular EDG6 genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that EDG6 genotype, haplotype or haplotype pair. Preferably, the EDG6 genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting EDG6 or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and an EDG6 genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population

and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the EDG6 gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and EDG6 genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their EDG6 genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the EDG6 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in PCT Application Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between EDG6 haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra,* Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in PCT Application Serial No. PCT/US00/17540.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the EDG6 gene. As described in PCT Application Serial No. PCT/US00/17540, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of EDG6 genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the EDG6 gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the EDG6 gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying EDG6 genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the EDG6 gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant EDG6 gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22 and PS23. Similarly, the nucleotide sequence of a variant fragment of the EDG6 gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the EDG6 gene, which is defined by haplotype 5, (or other reported EDG6 sequences) or to portions of the reference sequence (or other reported EDG6 sequences), except for genotyping oligonucleotides as described above.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of adenine at PS1, thymine at PS2, thymine at PS3, guanine at PS4, thymine at PS5, adenine at PS6, adenine at PS7, adenine at PS8, adenine at PS9, thymine at PS10, thymine at PS11, thymine at PS12, thymine at PS13, thymine at PS14, adenine at PS15, adenine at PS16, adenine at PS17, adenine at PS18, cytosine at PS19, adenine at PS20, adenine at PS21, thymine at PS22 and cytosine at PS23. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the EDG6 gene which is defined by any one of haplotypes 1- 4 and 6 - 24 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the EDG6 gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

EDG6 isogenes may be isolated using any method that allows separation of the two "copies" of the EDG6 gene present in an individual, which, as readily understood by the skilled artisan, may be

the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989, *supra*; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides EDG6 genome anthologies, which are collections of EDG6 isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. An EDG6 genome anthology may comprise individual EDG6 isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the EDG6 isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred EDG6 genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded EDG6 protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant EDG6 sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A

Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 *Science* 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the EDG6 gene will produce EDG6 mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of an EDG6 cDNA comprising a nucleotide sequence which is a polymorphic variant of the EDG6 reference coding sequence shown in Figure 2. Thus, the invention also provides EDG6 mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of thymine at a position corresponding to nucleotide 114, adenine at a position corresponding to nucleotide 231, adenine at a position corresponding to nucleotide 463, adenine at a position corresponding to nucleotide 490, adenine at a position corresponding to nucleotide 522, thymine at a position corresponding to nucleotide 565, thymine at a position corresponding to nucleotide 727, thymine at a position corresponding to nucleotide 804, thymine at a position corresponding to nucleotide 1059, thymine at a position corresponding to nucleotide 1094 and adenine at a position corresponding to nucleotide 1141. A particularly preferred polymorphic cDNA variant comprises the coding sequence of an EDG6 isogene defined by haplotypes 3c, 7c-12c, 19c-22c, and 24c. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized EDG6 cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of an EDG6 gene fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the EDG6 polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the EDG6 gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target

region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the EDG6 genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular EDG6 protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the EDG6 isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular EDG6 isogene. Expression of an EDG6 isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3′ untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions –10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of EDG6 mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of EDG6 mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2′ O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference EDG6 amino acid sequence shown in Figure 3. The location of a variant amino acid in an EDG6 polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig. 3). An EDG6 protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 except for having one or more variant amino acids selected from the group consisting of arginine at a position corresponding to amino acid position 155, serine at a position corresponding to amino acid position 164, serine at a position corresponding to amino acid position 189, cysteine at a position corresponding to amino acid position 243, leucine at a position corresponding to amino acid position 365 and methionine at a position corresponding to amino acid position 381. The invention specifically excludes amino acid sequences identical to those previously identified for EDG6,

including SEQ ID NO:3, and previously described fragments thereof. EDG6 protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, an EDG6 protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes shown in Table 5.

Table 2. Novel Polymorphic Variants of EDG6

Polymorphic Variant Number	Amino Acid Position and Identities					
	155	164	189	243	365	381
1	G	G	A	R	R	M
2	G	G	A	R	L	V
3	G	G	A	R	L	M
4	G	G	A	C	R	V
5	G	G	A	C	R	M
6	G	G	A	C	L	V
7	G	G	A	C	L	M
8	G	G	S	R	R	V
9	G	G	S	R	R	M
10	G	G	S	R	L	V
11	G	G	S	R	L	M
12	G	G	S	C	R	V
13	G	G	S	C	R	M
14	G	G	S	C	L	V
15	G	G	S	C	L	M
16	G	S	A	R	R	V
17	G	S	A	R	R	M
18	G	S	A	R	L	V
19	G	S	A	R	L	M
20	G	S	A	C	R	V
21	G	S	A	C	R	M
22	G	S	A	C	L	V
23	G	S	A	C	L	M
24	G	S	S	R	R	V
25	G	S	S	R	R	M
26	G	S	S	R	L	V
27	G	S	S	R	L	M
28	G	S	S	C	R	V
29	G	S	S	C	R	M
30	G	S	S	C	L	V
31	G	S	S	C	L	M
32	R	G	A	R	R	V

Table 2 Cont. Novel Polymorphic Variants of EDG6

Polymorphic Variant Number	Amino Acid Position and Identities					
	155	164	189	243	365	381
33	R	G	A	R	R	M
34	R	G	A	R	L	V
35	R	G	A	R	L	M
36	R	G	A	C	R	V
37	R	G	A	C	R	M
38	R	G	A	C	L	V
39	R	G	A	C	L	M
40	R	G	S	R	R	V
41	R	G	S	R	R	M
42	R	G	S	R	L	V
43	R	G	S	R	L	M
44	R	G	S	C	R	V
45	R	G	S	C	R	M
46	R	G	S	C	L	V
47	R	G	S	C	L	M
48	R	S	A	R	R	V
49	R	S	A	R	R	M
50	R	S	A	R	L	V
51	R	S	A	R	L	M
52	R	S	A	C	R	V
53	R	S	A	C	R	M
54	R	S	A	C	L	V
55	R	S	A	C	L	M
56	R	S	S	R	R	V
57	R	S	S	R	R	M
58	R	S	S	R	L	V
59	R	S	S	R	L	M
60	R	S	S	C	R	V
61	R	S	S	C	R	M
62	R	S	S	C	L	V
63	R	S	S	C	L	M

The invention also includes EDG6 peptide variants, which are any fragments of an EDG6 protein variant that contain one or more of the amino acid variations shown in Table 2 An EDG6 peptide variant is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such EDG6 peptide variants may be useful as antigens to generate antibodies specific for one of the above EDG6 isoforms. In addition, the EDG6 peptide variants may be useful in drug screening assays.

An EDG6 variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant EDG6 genomic and cDNA sequences as described above. Alternatively, the EDG6 protein variant may be isolated from a biological sample of an individual having an EDG6 isogene which encodes the variant protein. Where the sample contains two different EDG6 isoforms (i.e., the individual has different EDG6 isogenes), a particular EDG6 isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically

binds to that particular EDG6 isoform but does not bind to the other EDG6 isoform.

The expressed or isolated EDG6 protein may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the EDG6 protein as discussed further below. EDG6 variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant EDG6 gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric EDG6 protein. The non-EDG6 portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the EDG6 and non-EDG6 portions so that the EDG6 protein may be cleaved and purified away from the non-EDG6 portion.

An additional embodiment of the invention relates to using a novel EDG6 protein isoform in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known EDG6 protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The EDG6 protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to an EDG6 variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the EDG6 protein(s) of interest and then washed. Bound EDG6 protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel EDG6 protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the EDG6 protein.

In yet another embodiment, when a particular EDG6 haplotype or group of EDG6 haplotypes encodes an EDG6 protein variant with an amino acid sequence distinct from that of EDG6 protein isoforms encoded by other EDG6 haplotypes, then detection of that particular EDG6 haplotype or group of EDG6 haplotypes may be accomplished by detecting expression of the encoded EDG6 protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel EDG6 variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The EDG6 protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the EDG6 protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to

enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel protein isoforms described herein is administered to an individual to neutralize activity of the EDG6 isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the EDG6 protein variant from solution as well as react with EDG6 protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect EDG6 protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel EDG6 protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the EDG6 protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: Laboratory Techniques in Biochemistry and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The

antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci.USA 86;10029).

Effect(s) of the polymorphisms identified herein on expression of EDG6 may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the EDG6 gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into EDG6 protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired EDG6 isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the EDG6 isogene is introduced into a cell in such a way that it recombines with the endogenous EDG6 gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired EDG6 gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the EDG6 isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the EDG6 isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant EDG6 gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the EDG6 isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman

primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human EDG6 isogene and producing human EDG6 protein can be used as biological models for studying diseases related to abnormal EDG6 expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel EDG6 isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel EDG6 isogenes; an antisense oligonucleotide directed against one of the novel EDG6 isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel EDG6 isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel EDG6 isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the EDG6 gene and its genomic

variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The EDG6 polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the EDG6 gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the EDG6 gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1.

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	3484-3507	complement of 4101-4078	618 nt
Fragment 2	3698-3717	complement of 4427-4407	730 nt
Fragment 3	3899-3918	complement of 4509-4491	611 nt
Fragment 4	3938-3960	complement of 4639-4617	702 nt
Fragment 5	4266-4286	complement of 4941-4920	676 nt
Fragment 6	4625-4646	complement of 5188-5169	564 nt
Fragment 7	4915-4937	complement of 5583-5563	669 nt
Fragment 8	5070-5089	complement of 5771-5749	702 nt

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 µl
10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 µl
100 ng of human genomic DNA	= 1 µl
10 mM dNTP	= 0.4 µl
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 µl
Forward Primer (10 µM)	= 0.4 µl
Reverse Primer (10 µM)	= 0.4 µl
Water	= 6.6µl

Amplification profile:

97°C - 2 min. 1 cycle

97°C - 15 sec.
70°C - 45 sec. } 10 cycles
72°C - 45 sec.

97°C - 15 sec.
64°C - 45 sec. } 35 cycles
72°C - 45 sec.

Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 µl 384 well unifilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 µl of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were sequenced in both directions using the primer sets described previously or those represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

Fragment No.	Forward Primer	Reverse Primer
Fragment 1	3516-3534	complement of 4059-4041
Fragment 2	3757-3776	complement of 4321-4302
Fragment 3	3914-3932	complement of 4386-4368
Fragment 4	4071-4090	complement of 4608-4589
Fragment 5	4403-4423	complement of 4875-4857
Fragment 6	4670-4689	complement of 5152-5134
Fragment 7	4939-4958	complement of 5428-5410
Fragment 8	5149-5168	complement of 5691-5671

Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the EDG6 gene are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the EDG6 Gene

Polymorphic Site Number	PolyId[a]	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant Position	AA Variant
PS1	3216843	3591	G	A		
PS2	3216845	3697	C	T		
PS3	3216847	3804	C	T		
PS4	3216851	3818	A	G		
PS5	3216859	4123	C	T	114	R38R
PS6	3216861	4240	G	A	231	S77S
PS7	3216863	4472	G	A	463	G155R
PS8	3216865	4499	G	A	490	G164S
PS9	3216867	4531	G	A	522	A174A
PS10	3216869	4574	G	T	565	A189S
PS11	3216871	4736	C	T	727	R243C
PS12	3216873	4813	C	T	804	F268F
PS13	3216877	5068	C	T	1059	S353S
PS14	3216879	5103	G	T	1094	R365L
PS15	3216883	5150	G	A	1141	V381M
PS16	3216885	5179	G	A		
PS17	3216887	5301	G	A		
PS18	3216889	5333	G	A		
PS19	3216893	5448	G	C		
PS20	3216895	5560	G	A		
PS21	3216899	5580	G	A		
PS22	3216901	5587	C	T		
PS23	3216903	5606	G	C		

[a]PolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.

EXAMPLE 2

This example illustrates analysis of the EDG6 polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 4 (Part1). Genotypes and Haplotype Pairs Observed for EDG6 Gene

Genotype Number	Polymorphic Sites PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	HAP Pair
1	G	C	C	G	C	G	G	G	G	G	18 18
2	G	C	C	G	C	G	G	G	G	G	17 17
3	G	C	C	A	C	G	G	G	G	G	5 5
4	G	C	C	G	C	G	G	G	G	G	16 16
5	G	C/T	C	G/A	C	G	G	G	G	G	17 24
6	G	C	C	A	C	G	G	G	G	G	5 7
7	G	C	C	G	C	G	G/A	G	G	G	17 9
8	G	C	C	G	C	G	G	G	G	G	17 20
9	G	C	C	G	C/T	G	G	G	G	G	17 22
10	G/A	C	C	G	C	G	G	G	G	G	17 1
11	G	C	C	G	C	G	G	G/A	G	G	17 10
12	G	C	C	G/A	C	G	G	G	G	G	17 6
13	G	C	C	G	C	G	G	G	G	G	17 12
14	G	C	C	G/A	C	G	G	G	G	G	17 7
15	G	C	C	G	C	G	G	G	G	G	17 13
16	G	C	C	G	C	G	G	G/A	G	G	18 10
17	G	C	C	A	C	G/A	G	G	G	G	5 3
18	G	C	C	G/A	C	G/A	G	G	G	G	17 3
19	G	C	C	A	C	G	G	G	G	G	5 6
20	G	C	C	G	C	G/A	G	G	G	G	17 8
21	G	C	C	G	C	G	G	G	G	G	18 14
22	G	C	C	G	C	G	G	G	G	G	17 14
23	G	C	C	G/A	C	G	G	G	G	G	17 5
24	G	C	C	G	C	G	G	G	G	G	17 15
25	G	C	C/T	G	C	G	G	G	G	G	17 23
26	G	C	C	G	C	G	G	G	G/A	G	17 11
27	G	C	C	A	C	G	G	G	G	G	5 4
28	G	C	C	G/A	C	G	G	G	G	G	18 6
29	G	C	C	G	C	G	G	G	G	G	17 18
30	G	C	C	G	C	G	G	G	G	G	17 16
31	G	C	C	G/A	C	G/A	G	G	G	G	18 3
32	G	C	C	G	C	G	G	G	G	G/T	17 21
33	G	C	C	G/A	C	G	G	G	G	G	18 5
34	G/A	C	C	G	C	G	G	G	G	G	17 2
35	G	C	C	G	C	G	G	G	G	G	17 19

Table 4(Part2). Genotypes and Haplotype Pairs Observed for EDG6 Gene

Genotype Number	Polymorphic Sites PS11	PS12	PS13	PS14	PS15	PS16	PS17	PS18	PS19	PS20	HAP Pair	
1	C	C	C	G	G	G	G	G	G	G	18	18
2	C	C	C	G	G	G	G	G	G	G	17	17
3	C	C	C	G	G	G	G	G	G	G	5	5
4	C	C	C	G	G	G	G	G	G	G	16	16
5	C	C	C	G/T	G	G	G	G	G	G	17	24
6	C/T	C	C	G	G	G	G	G	G	G	5	7
7	C	C	C	G	G	G	G	G	G	G	17	9
8	C	C	C/T	G	G	G	G	G	G	G	17	20
9	C	C	C	G	G	G	G	G	G	G	17	22
10	C	C	C	G	G	G	G	G	G	G	17	1
11	C	C	C	G	G	G	G	G	G	G	17	10
12	C	C	C	G	G	G	G	G	G	G	17	6
13	C	C	C	G	G/A	G	G	G	G	G	17	12
14	C/T	C	C	G	G	G	G	G	G	G	17	7
15	C	C	C	G	G	G	G	G/A	G	G	17	13
16	C	C	C	G	G	G	G	G	G	G	18	10
17	C	C	C	G	G	G	G	G	G	G	5	3
18	C	C	C	G	G	G	G	G	G	G	17	3
19	C	C	C	G	G	G	G	G	G	G	5	6
20	C	C	C	G	G	G	G	G	G	G	17	8
21	C	C	C	G	G	G	G	G	G/C	G	18	14
22	C	C	C	G	G	G	G	G	G/C	G	17	14
23	C	C	C	G	G	G	G	G	G	G	17	5
24	C	C	C	G	G	G	G	G	G	G/A	17	15
25	C	C	C	G	G	G	G	G	G	G	17	23
26	C	C/T	C	G	G	G	G	G	G	G	17	11
27	C	C	C	G	G	G	G/A	G	G	G	5	4
28	C	C	C	G	G	G	G	G	G	G	18	6
29	C	C	C	G	G	G	G	G	G	G	17	18
30	C	C	C	G	G	G	G	G	G	G	17	16
31	C	C	C	G	G	G	G	G	G	G	18	3
32	C	C	C	G	G	G	G	G	G	G	17	21
33	C	C	C	G	G	G	G	G	G	G	18	5
34	C	C	C	G	G	G	G	G	G	G	17	2
35	C	C	C/T	G	G	G/A	G	G	G	G	17	19

Table 4(Part3). Genotypes and Haplotype Pairs Observed for EDG6 Gene

Genotype Number	Polymorphic Sites PS21	PS22	PS23	HAP	Pair
1	G	T	G	18	18
2	G	C	G	17	17
3	G	C	G	5	5
4	A	C	G	16	16
5	G	C	G	17	24
6	G	C/T	G	5	7
7	G	C/T	G	17	9
8	G	C	G	17	20
9	G	C	G	17	22
10	G	C	G	17	1
11	G	C	G	17	10
12	G	C/T	G	17	6
13	G	C	G	17	12
14	G	C/T	G	17	7
15	G	C	G	17	13

Table 4(Part3 cont.). Genotypes and Haplotype Pairs Observed for EDG6 Gene

Genotype Number	Polymorphic Sites PS21	PS22	PS23	HAP Pair	
16	G	T/C	G	18	10
17	G	C/T	G	5	3
18	G	C/T	G	17	3
19	G	C/T	G	5	6
20	G	C	G	17	8
21	G	T/C	G	18	14
22	G	C	G	17	14
23	G	C	G	17	5
24	G	C	G	17	15
25	G	C/T	G	17	23
26	G	C/T	G	17	11
27	G	C	G	5	4
28	G	T	G	18	6
29	G	C/T	G	17	18
30	G/A	C	G	17	16
31	G	T	G	18	3
32	G	C	G	17	21
33	G	T/C	G	18	5
34	G	C/T	G/C	17	2
35	G	C	G	17	19

The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample, as described in U.S. Provisional Application Serial No. 60/198,340 entitled "A Method and System for Determining Haplotypes from a Collection of Polymorphisms" and the corresponding International Application, PCT/US01/12831. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals. In our analysis, the list of haplotypes was augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family).

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 24 human EDG6 haplotypes shown in Table 5 below.

An EDG6 isogene defined by a full-haplotype shown in Table 5 below comprises the regions of the SEQ ID NOS indicated in Table 5, with their corresponding set of polymorphic locations and identities, which are also set forth in Table 5.

Table 5 (Part A). Haplotypes of the EDG6 Gene

Haplotype Number[a]										PS[b] No.	PS Position[c]	Seq ID No.[d]	Region Examined[e]
1	2	3	4	5	6	7	8	9	10				
A	A	G	G	G	G	G	G	G	G	1	3591	1/119	3484-5771
C	C	C	C	C	C	C	C	C	C	2	3697	1/119	3484-5771
C	C	C	C	C	C	C	C	C	C	3	3804	1/119	3484-5771
G	G	A	A	A	A	A	G	G	G	4	3818	1/119	3484-5771
C	C	C	C	C	C	C	C	C	C	5	4123	1/119	3484-5771
G	G	A	G	G	G	G	A	G	G	6	4240	1/119	3484-5771
G	G	G	G	G	G	G	G	A	G	7	4472	1/119	3484-5771
G	G	G	G	G	G	G	G	G	A	8	4499	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	9	4531	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	10	4574	1/119	3484-5771
C	C	C	C	C	C	T	C	C	C	11	4736	1/119	3484-5771
C	C	C	C	C	C	C	C	C	C	12	4813	1/119	3484-5771
C	C	C	C	C	C	C	C	C	C	13	5068	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	14	5103	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	15	5150	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	16	5179	1/119	3484-5771
G	G	G	A	G	G	G	G	G	G	17	5301	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	18	5333	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	19	5448	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	20	5560	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	21	5580	1/119	3484-5771
C	T	T	C	C	T	T	C	T	C	22	5587	1/119	3484-5771
G	C	G	G	G	G	G	G	G	G	23	5606	1/119	3484-5771

Table 5 (Part B). Haplotypes of the EDG6 Gene

Haplotype Number[a]										PS[b] No.	PS Pos.[c]	Seq ID No.[d]	Region Examined[e]
11	12	13	14	15	16	17	18	19	20				
G	G	G	G	G	G	G	G	G	G	1	3591	1/119	3484-5771
C	C	C	C	C	C	C	C	C	C	2	3697	1/119	3484-5771
C	C	C	C	C	C	C	C	C	C	3	3804	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	4	3818	1/119	3484-5771
C	C	C	C	C	C	C	C	C	C	5	4123	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	6	4240	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	7	4472	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	8	4499	1/119	3484-5771
A	G	G	G	G	G	G	G	G	G	9	4531	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	10	4574	1/119	3484-5771
C	C	C	C	C	C	C	C	C	C	11	4736	1/119	3484-5771
T	C	C	C	C	C	C	C	C	C	12	4813	1/119	3484-5771
C	C	C	C	C	C	C	C	T	T	13	5068	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	14	5103	1/119	3484-5771
G	A	G	G	G	G	G	G	G	G	15	5150	1/119	3484-5771
G	G	G	G	G	G	G	G	A	G	16	5179	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	17	5301	1/119	3484-5771
G	G	A	G	G	G	G	G	G	G	18	5333	1/119	3484-5771
G	G	G	C	G	G	G	G	G	G	19	5448	1/119	3484-5771
G	G	G	G	A	G	G	G	G	G	20	5560	1/119	3484-5771
G	G	G	G	G	A	G	G	G	G	21	5580	1/119	3484-5771
T	C	C	C	C	C	C	T	C	C	22	5587	1/119	3484-5771
G	G	G	G	G	G	G	G	G	G	23	5606	1/119	3484-5771

Table 5 (Part C). Haplotypes of the EDG6 Gene

Haplotype Number[a]				PS[b]	PS	Seq ID	Region
21	22	23	24	No.	Position[c]	No.[d]	Examined[e]
G	G	G	G	1	3591	1/119	3484-5771
C	C	C	T	2	3697	1/119	3484-5771
C	C	T	C	3	3804	1/119	3484-5771
G	G	G	A	4	3818	1/119	3484-5771
C	T	C	C	5	4123	1/119	3484-5771
G	G	G	G	6	4240	1/119	3484-5771
G	G	G	G	7	4472	1/119	3484-5771
G	G	G	G	8	4499	1/119	3484-5771
G	G	G	G	9	4531	1/119	3484-5771
T	G	G	G	10	4574	1/119	3484-5771
C	C	C	C	11	4736	1/119	3484-5771
C	C	C	C	12	4813	1/119	3484-5771
C	C	C	C	13	5068	1/119	3484-5771
G	G	G	T	14	5103	1/119	3484-5771
G	G	G	G	15	5150	1/119	3484-5771
G	G	G	G	16	5179	1/119	3484-5771
G	G	G	G	17	5301	1/119	3484-5771
G	G	G	G	18	5333	1/119	3484-5771
G	G	G	G	19	5448	1/119	3484-5771
G	G	G	G	20	5560	1/119	3484-5771
G	G	G	G	21	5580	1/119	3484-5771
C	C	T	C	22	5587	1/119	3484-5771
G	G	G	G	23	5606	1/119	3484-5771

[a]Alleles for EDG6 haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS within the indicated SEQ ID NO, with the 1st position number referring to the first SEQ ID NO and the 2nd position number referring to the 2nd SEQ ID NO;
[d]1st SEQ ID NO refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol; 2nd SEQ ID NO is a modified version of the 1st SEQ ID NO that comprises the context sequence of each polymorphic site, PS1-PS23, to facilitate electronic searching of the haplotypes;
[e]Region examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

SEQ ID NO:1 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol. SEQ ID NO:119 is a modified version of SEQ ID NO:1 that shows the context sequence of each of PS1-PS23 in a uniform format to facilitate electronic searching of the EDG6 haplotypes. For each polymorphic site, SEQ ID NO:119 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each polymorphic site is separated by genomic sequence whose composition is defined elsewhere herein.

Table 6 below shows the percent of chromosomes characterized by a given EDG6 haplotype for all unrelated individuals in the Index Repository for which haplotype data was obtained. The percent of these unrelated individuals who have a given EDG6 haplotype pair is shown in Table 7. In Tables 6 and 7, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according

to their self-identified ethnogeographic origin. Abbreviations used in Tables 6 and 7 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and AM = Native American.

Table 6. Frequency of Observed EDG6 Haplotypes In Unrelated Individuals

HAP No.	HAP ID	Total	CA	AF	AS	HL	AM
1	3219666	0.61	2.38	0.0	0.0	0.0	0.0
2	3219662	0.61	0.0	0.0	2.5	0.0	0.0
3	3219649	2.44	4.76	0.0	0.0	5.56	0.0
4	3219663	0.61	2.38	0.0	0.0	0.0	0.0
5	3219646	12.8	19.05	7.5	0.0	22.22	33.33
6	3219647	3.05	9.52	0.0	0.0	2.78	0.0
7	3219652	1.22	2.38	0.0	0.0	2.78	0.0
8	3219653	1.22	4.76	0.0	0.0	0.0	0.0
9	3219665	0.61	0.0	0.0	0.0	0.0	16.67
10	3219654	1.22	0.0	0.0	5.0	0.0	0.0
11	3219661	0.61	0.0	0.0	2.5	0.0	0.0
12	3219669	0.61	0.0	0.0	0.0	2.78	0.0
13	3219673	0.61	0.0	2.5	0.0	0.0	0.0
14	3219650	2.44	0.0	10.0	0.0	0.0	0.0
15	3219677	0.61	0.0	0.0	2.5	0.0	0.0
16	3219648	3.66	0.0	0.0	15.0	0.0	0.0
17	3219644	41.46	30.95	50.0	47.5	36.11	50.0
18	3219645	21.95	23.81	22.5	20.0	25.0	0.0
19	3219655	0.61	0.0	2.5	0.0	0.0	0.0
20	3219678	0.61	0.0	2.5	0.0	0.0	0.0
21	3219676	0.61	0.0	0.0	2.5	0.0	0.0
22	3219675	0.61	0.0	0.0	2.5	0.0	0.0
23	3219664	0.61	0.0	2.5	0.0	0.0	0.0
24	3219660	0.61	0.0	0.0	0.0	2.78	0.0

Table 7. Frequency of Observed EDG6 Haplotype Pairs In Unrelated Individuals

HAP1	HAP2	Total	CA	AF	AS	HL	AM
18	18	8.54	14.29	5.0	10.0	5.56	0.0
17	17	13.41	9.52	20.0	15.0	5.56	33.33
5	5	3.66	4.76	0.0	0.0	5.56	33.33
16	16	1.22	0.0	0.0	5.0	0.0	0.0
17	24	1.22	0.0	0.0	0.0	5.56	0.0
5	7	1.22	0.0	0.0	0.0	5.56	0.0
17	9	1.22	0.0	0.0	0.0	0.0	33.33
17	20	1.22	0.0	5.0	0.0	0.0	0.0
17	22	1.22	0.0	0.0	5.0	0.0	0.0
17	1	1.22	4.76	0.0	0.0	0.0	0.0
17	10	1.22	0.0	0.0	5.0	0.0	0.0
17	6	1.22	4.76	0.0	0.0	0.0	0.0
17	12	1.22	0.0	0.0	0.0	5.56	0.0
17	7	1.22	4.76	0.0	0.0	0.0	0.0
17	13	1.22	0.0	5.0	0.0	0.0	0.0
18	10	1.22	0.0	0.0	5.0	0.0	0.0
5	3	1.22	0.0	0.0	0.0	5.56	0.0
17	3	2.44	4.76	0.0	0.0	5.56	0.0
5	6	3.66	9.52	0.0	0.0	5.56	0.0
17	8	2.44	9.52	0.0	0.0	0.0	0.0
18	14	2.44	0.0	10.0	0.0	0.0	0.0
17	14	2.44	0.0	10.0	0.0	0.0	0.0
17	5	7.32	9.52	10.0	0.0	11.11	0.0
17	15	1.22	0.0	0.0	5.0	0.0	0.0
17	23	1.22	0.0	5.0	0.0	0.0	0.0
17	11	1.22	0.0	0.0	5.0	0.0	0.0
5	4	1.22	4.76	0.0	0.0	0.0	0.0
18	6	1.22	4.76	0.0	0.0	0.0	0.0
17	18	17.07	4.76	20.0	15.0	33.33	0.0
17	16	4.88	0.0	0.0	20.0	0.0	0.0
18	3	1.22	4.76	0.0	0.0	0.0	0.0
17	21	1.22	0.0	0.0	5.0	0.0	0.0
18	5	3.66	4.76	5.0	0.0	5.56	0.0
17	2	1.22	0.0	0.0	5.0	0.0	0.0
17	19	1.22	0.0	5.0	0.0	0.0	0.0

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5% in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from

that population group. In addition, the size and composition of the Index Repository means that the relative frequencies determined therein for the haplotypes and haplotype pairs of the EDG6 gene are likely to be similar to the relative frequencies of these EDG6 haplotypes and haplotype pairs in the general U.S. population and in the four population groups represented in the Index Repository. The genetic diversity observed for the three Native Americans is presented because it is of scientific interest, but due to the small sample size it lacks statistical significance.

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the endothelial differentiation, G-protein-coupled receptor 6 (EDG6) gene of an individual, which comprises determining which of the EDG6 haplotypes shown in the table immediately below defines one copy of the individual's EDG6 gene, wherein each of the EDG6 haplotypes comprises a set of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Number[a]										PS[b]	PS
1	2	3	4	5	6	7	8	9	10	Number	Position[c]
A	A	G	G	G	G	G	G	G	G	1	3591
C	C	C	C	C	C	C	C	C	C	2	3697
C	C	C	C	C	C	C	C	C	C	3	3804
G	G	A	A	A	A	A	G	G	G	4	3818
C	C	C	C	C	C	C	C	C	C	5	4123
G	G	A	G	G	G	G	A	G	G	6	4240
G	G	G	G	G	G	G	G	A	G	7	4472
G	G	G	G	G	G	G	G	G	A	8	4499
G	G	G	G	G	G	G	G	G	G	9	4531
G	G	G	G	G	G	G	G	G	G	10	4574
C	C	C	C	C	C	T	C	C	C	11	4736
C	C	C	C	C	C	C	C	C	C	12	4813
C	C	C	C	C	C	C	C	C	C	13	5068
G	G	G	G	G	G	G	G	G	G	14	5103
G	G	G	G	G	G	G	G	G	G	15	5150
G	G	G	G	G	G	G	G	G	G	16	5179
G	G	G	A	G	G	G	G	G	G	17	5301
G	G	G	G	G	G	G	G	G	G	18	5333
G	G	G	G	G	G	G	G	G	G	19	5448
G	G	G	G	G	G	G	G	G	G	20	5560
G	G	G	G	G	G	G	G	G	G	21	5580
C	T	T	C	C	T	T	C	T	C	22	5587
G	C	G	G	G	G	G	G	G	G	23	5606

Haplotype Number[a]										PS[b] Number	PS Position[c]
11	12	13	14	15	16	17	18	19	20		
G	G	G	G	G	G	G	G	G	G	1	3591
C	C	C	C	C	C	C	C	C	C	2	3697
C	C	C	C	C	C	C	C	C	C	3	3804
G	G	G	G	G	G	G	G	G	G	4	3818
C	C	C	C	C	C	C	C	C	C	5	4123
G	G	G	G	G	G	G	G	G	G	6	4240
G	G	G	G	G	G	G	G	G	G	7	4472
G	G	G	G	G	G	G	G	G	G	8	4499
A	G	G	G	G	G	G	G	G	G	9	4531
G	G	G	G	G	G	G	G	G	G	10	4574
C	C	C	C	C	C	C	C	C	C	11	4736
T	C	C	C	C	C	C	C	C	C	12	4813
C	C	C	C	C	C	C	C	T	T	13	5068
G	G	G	G	G	G	G	G	G	G	14	5103
G	A	G	G	G	G	G	G	G	G	15	5150
G	G	G	G	G	G	G	G	A	G	16	5179
G	G	G	G	G	G	G	G	G	G	17	5301
G	G	A	G	G	G	G	G	G	G	18	5333
G	G	G	C	G	G	G	G	G	G	19	5448
G	G	G	G	A	G	G	G	G	G	20	5560
G	G	G	G	G	A	G	G	G	G	21	5580
T	C	C	C	C	C	C	T	C	C	22	5587
G	G	G	G	G	G	G	G	G	G	23	5606

Haplotype Number[a]				PS[b] Number	PS Position[c]
21	22	23	24		
G	G	G	G	1	3591
C	C	C	T	2	3697
C	C	T	C	3	3804
G	G	G	A	4	3818
C	T	C	C	5	4123
G	G	G	G	6	4240
G	G	G	G	7	4472
G	G	G	G	8	4499
G	G	G	G	9	4531
T	G	G	G	10	4574
C	C	C	C	11	4736
C	C	C	C	12	4813
C	C	C	C	13	5068
G	G	G	T	14	5103
G	G	G	G	15	5150
G	G	G	G	16	5179
G	G	G	G	17	5301
G	G	G	G	18	5333
G	G	G	G	19	5448
G	G	G	G	20	5560
G	G	G	G	21	5580
C	C	T	C	22	5587
G	G	G	G	23	5606

[a]Alleles for haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS within SEQ ID NO:1.

2. The method of claim 1, wherein the determining step comprises identifying the phased

sequence of nucleotides present at each of PS1-PS23 on the one copy of the individual's EDG6 gene.

3. A method for haplotyping the endothelial differentiation, G-protein-coupled receptor 6 (EDG6) gene of an individual, which comprises determining which of the EDG6 haplotype pairs shown in the table immediately below defines both copies of the individual's EDG6 gene, wherein each of the EDG6 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Pair[a]								PS[b]	PS
18/18	17/17	5/5	16/16	17/24	5/7	17/9	17/20	Number	Position[c]
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	A/A	G/G	G/A	A/A	G/G	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	13	5068
G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	A/A	G/G	G/G	G/G	G/G	21	5580
T/T	C/C	C/C	C/C	C/C	C/T	C/T	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b]	PS
17/22	17/1	17/10	17/6	17/12	17/7	17/13	18/10	Number	Position[c]
G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	G/G	G/A	G/G	G/A	G/G	G/G	4	3818
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/A	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	21	5580
C/C	C/C	C/C	C/T	C/C	C/T	C/C	T/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b]	PS
5/3	17/3	5/6	17/8	18/14	17/14	17/5	17/15	Number	Position[c]
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
A/A	G/A	A/A	G/G	G/G	G/G	G/A	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/A	G/A	G/G	G/A	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/C	G/C	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	20	5560
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	21	5580
C/T	C/T	C/T	C/C	T/C	C/C	C/C	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b] Number	PS Position[c]
17/23	17/11	5/4	18/6	17/18	17/16	18/3	17/21		
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	A/A	G/A	G/G	G/G	G/A	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/T	10	4574
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	11	4736
C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G	21	5580
C/T	C/T	C/C	T/T	C/T	C/C	T/T	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]			PS[b] Number	PS Position[c]
18/5	17/2	17/19		
G/G	G/A	G/G	1	3591
C/C	C/C	C/C	2	3697
C/C	C/C	C/C	3	3804
G/A	G/G	G/G	4	3818
C/C	C/C	C/C	5	4123
G/G	G/G	G/G	6	4240
G/G	G/G	G/G	7	4472
G/G	G/G	G/G	8	4499
G/G	G/G	G/G	9	4531
G/G	G/G	G/G	10	4574
C/C	C/C	C/C	11	4736
C/C	C/C	C/C	12	4813
C/C	C/C	C/T	13	5068
G/G	G/G	G/G	14	5103
G/G	G/G	G/G	15	5150
G/G	G/G	G/A	16	5179
G/G	G/G	G/G	17	5301
G/G	G/G	G/G	18	5333
G/G	G/G	G/G	19	5448
G/G	G/G	G/G	20	5560
G/G	G/G	G/G	21	5580
T/C	C/T	C/C	22	5587
G/G	G/C	G/G	23	5606

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1.

4. The method of claim 3, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS23 on both copies of the individual's EDG6 gene.

5. A method for genotyping the endothelial differentiation, G-protein-coupled receptor 6 (EDG6) gene of an individual, comprising determining for the two copies of the EDG6 gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22 and PS23, wherein the one or more PS have the location and alternative alleles shown in SEQ ID NO:1.

6. The method of claim 5, wherein the determining step comprises:
   (a) isolating from the individual a nucleic acid mixture comprising both copies of the EDG6 gene, or a fragment thereof, that are present in the individual;
   (b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;
   (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
   (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
   (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

7. The method of claim 5, which comprises determining for the two copies of the EDG6 gene present in the individual the identity of the nucleotide pair at each of PS1-PS23.

8. A method for haplotyping the endothelial differentiation, G-protein-coupled receptor 6 (EDG6) gene of an individual which comprises determining, for one copy of the EDG6 gene present in the individual, the identity of the nucleotide at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22 and PS23, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:1.

9. The method of claim 8, wherein the determining step comprises:
   (a) isolating from the individual a nucleic acid sample containing only one of the two copies of the EDG6 gene, or a fragment thereof, that is present in the individual;
   (b) amplifying from the nucleic acid sample a target region containing the selected polymorphic site;
   (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
   (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides

present at the selected polymorphic site; and

(e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

10. A method for predicting a haplotype pair for the endothelial differentiation, G-protein-coupled receptor 6 (EDG6) gene of an individual comprising:

(a) identifying an EDG6 genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22 and PS23, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:1;

(b) enumerating all possible haplotype pairs which are consistent with the genotype;

(c) comparing the possible haplotype pairs to the haplotype pair data set forth in the table immediately below; and

(d) assigning a haplotype pair to the individual that is consistent with the data

Haplotype Pair[a]								PS[b]	PS
18/18	17/17	5/5	16/16	17/24	5/7	17/9	17/20	Number	Position[c]
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	A/A	G/G	G/A	A/A	G/G	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	13	5068
G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	A/A	G/G	G/G	G/G	G/G	21	5580
T/T	C/C	C/C	C/C	C/C	C/T	C/T	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b]	PS
17/22	17/1	17/10	17/6	17/12	17/7	17/13	18/10	Number	Position[c]
G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	G/G	G/A	G/G	G/A	G/G	G/G	4	3818
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/A	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	21	5580
C/C	C/C	C/C	C/T	C/C	C/T	C/C	T/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b]	PS
5/3	17/3	5/6	17/8	18/14	17/14	17/5	17/15	Number	Position[c]
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
A/A	G/A	A/A	G/G	G/G	G/G	G/A	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/A	G/A	G/G	G/A	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/C	G/C	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	20	5560
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	21	5580
C/T	C/T	C/T	C/C	T/C	C/C	C/C	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b] Number	PS Position[c]
17/23	17/11	5/4	18/6	17/18	17/16	18/3	17/21		
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	A/A	G/A	G/G	G/G	G/A	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/T	10	4574
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	11	4736
C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G	21	5580
C/T	C/T	C/C	T/T	C/T	C/C	T/T	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]			PS[b] Number	PS Position[c]
18/5	17/2	17/19		
G/G	G/A	G/G	1	3591
C/C	C/C	C/C	2	3697
C/C	C/C	C/C	3	3804
G/A	G/G	G/G	4	3818
C/C	C/C	C/C	5	4123
G/G	G/G	G/G	6	4240
G/G	G/G	G/G	7	4472
G/G	G/G	G/G	8	4499
G/G	G/G	G/G	9	4531
G/G	G/G	G/G	10	4574
C/C	C/C	C/C	11	4736
C/C	C/C	C/C	12	4813
C/C	C/C	C/T	13	5068
G/G	G/G	G/G	14	5103
G/G	G/G	G/G	15	5150
G/G	G/G	G/A	16	5179
G/G	G/G	G/G	17	5301
G/G	G/G	G/G	18	5333
G/G	G/G	G/G	19	5448
G/G	G/G	G/G	20	5560
G/G	G/G	G/G	21	5580
T/C	C/T	C/C	22	5587
G/G	G/C	G/G	23	5606

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1.

11. The method of claim 10, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS23, which have the location and alternative alleles shown in SEQ ID NO:1.

12. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the endothelial differentiation, G-protein-coupled receptor 6 (EDG6) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-24 shown in the table presented immediately below, wherein each of the haplotypes comprises a set of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Number[a]										PS[b]	PS
1	2	3	4	5	6	7	8	9	10	Number	Position[c]
A	A	G	G	G	G	G	G	G	G	1	3591
C	C	C	C	C	C	C	C	C	C	2	3697
C	C	C	C	C	C	C	C	C	C	3	3804
G	G	A	A	A	A	A	G	G	G	4	3818
C	C	C	C	C	C	C	C	C	C	5	4123
G	G	A	G	G	G	G	A	G	G	6	4240
G	G	G	G	G	G	G	G	A	G	7	4472
G	G	G	G	G	G	G	G	G	A	8	4499
G	G	G	G	G	G	G	G	G	G	9	4531
G	G	G	G	G	G	G	G	G	G	10	4574
C	C	C	C	C	C	T	C	C	C	11	4736
C	C	C	C	C	C	C	C	C	C	12	4813
C	C	C	C	C	C	C	C	C	C	13	5068
G	G	G	G	G	G	G	G	G	G	14	5103
G	G	G	G	G	G	G	G	G	G	15	5150
G	G	G	G	G	G	G	G	G	G	16	5179
G	G	G	A	G	G	G	G	G	G	17	5301
G	G	G	G	G	G	G	G	G	G	18	5333
G	G	G	G	G	G	G	G	G	G	19	5448
G	G	G	G	G	G	G	G	G	G	20	5560
G	G	G	G	G	G	G	G	G	G	21	5580
C	T	T	C	C	T	T	C	T	C	22	5587
G	C	G	G	G	G	G	G	G	G	23	5606

Haplotype Number[a]										PS[b] Number	PS Position[c]
11	12	13	14	15	16	17	18	19	20		
G	G	G	G	G	G	G	G	G	G	1	3591
C	C	C	C	C	C	C	C	C	C	2	3697
C	C	C	C	C	C	C	C	C	C	3	3804
G	G	G	G	G	G	G	G	G	G	4	3818
C	C	C	C	C	C	C	C	C	C	5	4123
G	G	G	G	G	G	G	G	G	G	6	4240
G	G	G	G	G	G	G	G	G	G	7	4472
G	G	G	G	G	G	G	G	G	G	8	4499
A	G	G	G	G	G	G	G	G	G	9	4531
G	G	G	G	G	G	G	G	G	G	10	4574
C	C	C	C	C	C	C	C	C	C	11	4736
T	C	C	C	C	C	C	C	C	C	12	4813
C	C	C	C	C	C	C	C	T	T	13	5068
G	G	G	G	G	G	G	G	G	G	14	5103
G	A	G	G	G	G	G	G	G	G	15	5150
G	G	G	G	G	G	G	G	A	G	16	5179
G	G	G	G	G	G	G	G	G	G	17	5301
G	G	A	G	G	G	G	G	G	G	18	5333
G	G	G	C	G	G	G	G	G	G	19	5448
G	G	G	G	A	G	G	G	G	G	20	5560
G	G	G	G	G	A	G	G	G	G	21	5580
T	C	C	C	C	C	C	T	C	C	22	5587
G	G	G	G	G	G	G	G	G	G	23	5606

Haplotype Number[a]				PS[b] Number	PS Position[c]
21	22	23	24		
G	G	G	G	1	3591
C	C	C	T	2	3697
C	C	T	C	3	3804
G	G	G	A	4	3818
C	T	C	C	5	4123
G	G	G	G	6	4240
G	G	G	G	7	4472
G	G	G	G	8	4499
G	G	G	G	9	4531
T	G	G	G	10	4574
C	C	C	C	11	4736
C	C	C	C	12	4813
C	C	C	C	13	5068
G	G	G	T	14	5103
G	G	G	G	15	5150
G	G	G	G	16	5179
G	G	G	G	17	5301
G	G	G	G	18	5333
G	G	G	G	19	5448
G	G	G	G	20	5560
G	G	G	G	21	5580
C	C	T	C	22	5587
G	G	G	G	23	5606

[a]Alleles for haplotypes are presented 5′ to 3′ in each column

[b]PS = polymorphic site;

[c]Position of PS in SEQ ID NO:1;

and wherein the haplotype pair is selected from the haplotype pairs shown in the table

immediately below, wherein each of the EDG6 haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Pair[a]								PS[b] Number	PS Position[c]
18/18	17/17	5/5	16/16	17/24	5/7	17/9	17/20		
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	A/A	G/G	G/A	A/A	G/G	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	13	5068
G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	A/A	G/G	G/G	G/G	G/G	21	5580
T/T	C/C	C/C	C/C	C/C	C/T	C/T	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b] Number	PS Position[c]
17/22	17/1	17/10	17/6	17/12	17/7	17/13	18/10		
G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	G/G	G/A	G/G	G/A	G/G	G/G	4	3818
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/A	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	21	5580
C/C	C/C	C/C	C/T	C/C	C/T	C/C	T/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b]	PS
5/3	17/3	5/6	17/8	18/14	17/14	17/5	17/15	Number	Position[c]
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
A/A	G/A	A/A	G/G	G/G	G/G	G/A	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/A	G/A	G/G	G/A	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/C	G/C	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	20	5560
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	21	5580
C/T	C/T	C/T	C/C	T/C	C/C	C/C	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b]	PS
17/23	17/11	5/4	18/6	17/18	17/16	18/3	17/21	Number	Position[c]
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	A/A	G/A	G/G	G/G	G/A	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/T	10	4574
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	11	4736
C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G	21	5580
C/T	C/T	C/C	T/T	C/T	C/C	T/T	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]			PS[b]	PS
18/5	17/2	17/19	Number	Position[c]
G/G	G/A	G/G	1	3591
C/C	C/C	C/C	2	3697
C/C	C/C	C/C	3	3804
G/A	G/G	G/G	4	3818
C/C	C/C	C/C	5	4123
G/G	G/G	G/G	6	4240
G/G	G/G	G/G	7	4472
G/G	G/G	G/G	8	4499
G/G	G/G	G/G	9	4531
G/G	G/G	G/G	10	4574
C/C	C/C	C/C	11	4736
C/C	C/C	C/C	12	4813
C/C	C/C	C/T	13	5068
G/G	G/G	G/G	14	5103
G/G	G/G	G/G	15	5150
G/G	G/G	G/A	16	5179
G/G	G/G	G/G	17	5301
G/G	G/G	G/G	18	5333
G/G	G/G	G/G	19	5448
G/G	G/G	G/G	20	5560
G/G	G/G	G/G	21	5580
T/C	C/T	C/C	22	5587
G/G	G/C	G/G	23	5606

[a]Haplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;

wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

13. The method of claim 12, wherein the trait is a clinical response to a drug targeting EDG6.

14. An isolated genotyping oligonucleotide for detecting a polymorphism in the endothelial differentiation, G-protein-coupled receptor 6 (EDG6) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22 and PS23, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:1.

15. The isolated genotyping oligonucleotide of claim 14, which is an allele-specific oligonucleotide that specifically hybridizes to an allele of the EDG6 gene at a region containing the polymorphic site.

16. The allele-specific oligonucleotide of claim 15, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-26, the complements of SEQ ID NOS:4-26, and SEQ ID NOS:27-72.

17. The isolated genotyping oligonucleotide of claim 14, which is a primer-extension oligonucleotide.

18. The primer-extension oligonucleotide of claim 17,which comprises a nucleotide sequence

selected from the group consisting of SEQ ID NOS:73-118.

19. A kit for genotyping the endothelial differentiation, G-protein-coupled receptor 6 (EDG6) gene of an individual, which comprises a set of oligonucleotides designed to genotype each of polymorphic sites (PS) PS1, PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS9, PS10, PS11, PS12, PS13, PS14, PS15, PS16, PS17, PS18, PS19, PS20, PS21, PS22 and PS23, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:1.

20. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a first nucleotide sequence which comprises a endothelial differentiation, G-protein-coupled receptor 6 (EDG6) isogene, wherein the EDG6 isogene is selected from the group consisting of isogenes 1- 4 and 6 - 24 shown in the table immediately below and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1- 4 and 6 - 24 is further defined by the corresponding set of polymorphisms whose locations and identities are set forth in the table immediately below

Isogene Number[a]										PS[b]	PS	SEQ ID	Region
1	2	3	4	5	6	7	8	9	10	No.	Pos.[c]	No.	Examined[d]
A	A	G	G	G	G	G	G	G	G	1	3591	1	3484-5771
C	C	C	C	C	C	C	C	C	C	2	3697	1	3484-5771
C	C	C	C	C	C	C	C	C	C	3	3804	1	3484-5771
G	G	A	A	A	A	A	G	G	G	4	3818	1	3484-5771
C	C	C	C	C	C	C	C	C	C	5	4123	1	3484-5771
G	G	A	G	G	G	G	A	G	G	6	4240	1	3484-5771
G	G	G	G	G	G	G	G	A	G	7	4472	1	3484-5771
G	G	G	G	G	G	G	G	G	A	8	4499	1	3484-5771
G	G	G	G	G	G	G	G	G	G	9	4531	1	3484-5771
G	G	G	G	G	G	G	G	G	G	10	4574	1	3484-5771
C	C	C	C	C	C	T	C	C	C	11	4736	1	3484-5771
C	C	C	C	C	C	C	C	C	C	12	4813	1	3484-5771
C	C	C	C	C	C	C	C	C	C	13	5068	1	3484-5771
G	G	G	G	G	G	G	G	G	G	14	5103	1	3484-5771
G	G	G	G	G	G	G	G	G	G	15	5150	1	3484-5771
G	G	G	G	G	G	G	G	G	G	16	5179	1	3484-5771
G	G	G	A	G	G	G	G	G	G	17	5301	1	3484-5771
G	G	G	G	G	G	G	G	G	G	18	5333	1	3484-5771
G	G	G	G	G	G	G	G	G	G	19	5448	1	3484-5771
G	G	G	G	G	G	G	G	G	G	20	5560	1	3484-5771
G	G	G	G	G	G	G	G	G	G	21	5580	1	3484-5771
C	T	T	C	C	T	T	C	T	C	22	5587	1	3484-5771
G	C	G	G	G	G	G	G	G	G	23	5606	1	3484-5771

Isogene Number[a]										PS[b] No.	PS Pos.[c]	SEQ ID No.	Region Examined[d]
11	12	13	14	15	16	17	18	19	20				
G	G	G	G	G	G	G	G	G	G	1	3591	1	3484-5771
C	C	C	C	C	C	C	C	C	C	2	3697	1	3484-5771
C	C	C	C	C	C	C	C	C	C	3	3804	1	3484-5771
G	G	G	G	G	G	G	G	G	G	4	3818	1	3484-5771
C	C	C	C	C	C	C	C	C	C	5	4123	1	3484-5771
G	G	G	G	G	G	G	G	G	G	6	4240	1	3484-5771
G	G	G	G	G	G	G	G	G	G	7	4472	1	3484-5771
G	G	G	G	G	G	G	G	G	G	8	4499	1	3484-5771
A	G	G	G	G	G	G	G	G	G	9	4531	1	3484-5771
G	G	G	G	G	G	G	G	G	G	10	4574	1	3484-5771
C	C	C	C	C	C	C	C	C	C	11	4736	1	3484-5771
T	C	C	C	C	C	C	C	C	C	12	4813	1	3484-5771
C	C	C	C	C	C	C	C	T	T	13	5068	1	3484-5771
G	G	G	G	G	G	G	G	G	G	14	5103	1	3484-5771
G	A	G	G	G	G	G	G	G	G	15	5150	1	3484-5771
G	G	G	G	G	G	G	G	A	G	16	5179	1	3484-5771
G	G	G	G	G	G	G	G	G	G	17	5301	1	3484-5771
G	G	A	G	G	G	G	G	G	G	18	5333	1	3484-5771
G	G	G	C	G	G	G	G	G	G	19	5448	1	3484-5771
G	G	G	G	A	G	G	G	G	G	20	5560	1	3484-5771
G	G	G	G	G	A	G	G	G	G	21	5580	1	3484-5771
T	C	C	C	C	C	C	T	C	C	22	5587	1	3484-5771
G	G	G	G	G	G	G	G	G	G	23	5606	1	3484-5771

Isogene Number[a]				PS[b] No.	PS Pos.[c]	SEQ ID No.	Region Examined[d]
21	22	23	24				
G	G	G	G	1	3591	1	3484-5771
C	C	C	T	2	3697	1	3484-5771
C	C	T	C	3	3804	1	3484-5771
G	G	G	A	4	3818	1	3484-5771
C	T	C	C	5	4123	1	3484-5771
G	G	G	G	6	4240	1	3484-5771
G	G	G	G	7	4472	1	3484-5771
G	G	G	G	8	4499	1	3484-5771
G	G	G	G	9	4531	1	3484-5771
T	G	G	G	10	4574	1	3484-5771
C	C	C	C	11	4736	1	3484-5771
C	C	C	C	12	4813	1	3484-5771
C	C	C	C	13	5068	1	3484-5771
G	G	G	T	14	5103	1	3484-5771
G	G	G	G	15	5150	1	3484-5771
G	G	G	G	16	5179	1	3484-5771
G	G	G	G	17	5301	1	3484-5771
G	G	G	G	18	5333	1	3484-5771
G	G	G	G	19	5448	1	3484-5771
G	G	G	G	20	5560	1	3484-5771
G	G	G	G	21	5580	1	3484-5771
C	C	T	C	22	5587	1	3484-5771
G	G	G	G	23	5606	1	3484-5771

[a]Alleles for isogenes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;
[d]Region examined represents the nucleotide positions defining the start and stop positions within the SEQ ID NO of the sequenced region.

(b) a second nucleotide sequence which comprises a fragment of the first nucleotide sequence, wherein the fragment comprises one or more polymorphisms selected from the group consisting of adenine at PS1, thymine at PS2, thymine at PS3, guanine at PS4, thymine at PS5, adenine at PS6, adenine at PS7, adenine at PS8, adenine at PS9, thymine at PS10, thymine at PS11, thymine at PS12, thymine at PS13, thymine at PS14, adenine at PS15, adenine at PS16, adenine at PS17, adenine at PS18, cytosine at PS19, adenine at PS20, adenine at PS21, thymine at PS22 and cytosine at PS23, wherein the selected polymorphism has the location set forth in the table immediately above; and

(c) a third nucleotide sequence which is complementary to the first or second nucleotide sequence.

21. The isolated polynucleotide of claim 20, which is a DNA molecule and comprises both the first and third nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.

22. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 20, wherein the organism expresses an EDG6 protein encoded by the first nucleotide sequence.

23. The recombinant nonhuman organism of claim 22, which is a transgenic animal.

24. The isolated polynucleotide of claim 20 which consists of the second nucleotide sequence.

25. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a coding sequence for a endothelial differentiation, G-protein-coupled receptor 6 (EDG6) isogene wherein the coding sequence is defined by a haplotype selected from the group consisting of 3c, 7c-12c, 19c-22c, and 24c shown in the table immediately below and wherein the coding sequence comprises SEQ ID NO:2 except at each of the polymorphic sites which have the locations and polymorphisms set forth in the table immediately below:

Coding Sequence Haplotype Number[a]												PS[b]	PS
3c	7c	8c	9c	10c	11c	12c	19c	20c	21c	22c	24c	No.	Position[c]
C	C	C	C	C	C	C	C	C	C	T	C	5	114
A	G	A	G	G	G	G	G	G	G	G	G	6	231
G	G	G	A	G	G	G	G	G	G	G	G	7	463
G	G	G	G	A	G	G	G	G	G	G	G	8	490
G	G	G	G	G	A	G	G	G	G	G	G	9	522
G	G	G	G	G	G	G	G	G	T	G	G	10	565
C	T	C	C	C	C	C	C	C	C	C	C	11	727
C	C	C	C	C	T	C	C	C	C	C	C	12	804
C	C	C	C	C	C	C	T	T	C	C	C	13	1059
G	G	G	G	G	G	G	G	G	G	G	T	14	1094
G	G	G	G	G	G	A	G	G	G	G	G	15	1141

[a]Alleles for coding sequence haplotypes are presented 5′ to 3′ in each column; the numerical portion of the coding sequence haplotype number represents the number of the

parent full EDG6 haplotype;
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:2;

and

(b) a fragment of the coding sequence, wherein the fragment comprises at least one polymorphism selected from the group consisting of thymine at a position corresponding to nucleotide 114, adenine at a position corresponding to nucleotide 231, adenine at a position corresponding to nucleotide 463, adenine at a position corresponding to nucleotide 490, adenine at a position corresponding to nucleotide 522, thymine at a position corresponding to nucleotide 565, thymine at a position corresponding to nucleotide 727, thymine at a position corresponding to nucleotide 804, thymine at a position corresponding to nucleotide 1059, thymine at a position corresponding to nucleotide 1094 and adenine at a position corresponding to nucleotide 1141, wherein said positions in the coding sequence and the fragment refer to SEQ ID NO:2.

26. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 25, wherein the organism expresses a endothelial differentiation, G-protein-coupled receptor 6 (EDG6) protein encoded by the polymorphic variant sequence.

27. The recombinant nonhuman organism of claim 26, which is a transgenic animal.

28. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the endothelial differentiation, G-protein-coupled receptor 6 (EDG6) protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:3 and the polymorphic variant comprises one or more variant amino acids selected from the group consisting of arginine at a position corresponding to amino acid position 155, serine at a position corresponding to amino acid position 164, serine at a position corresponding to amino acid position 189, cysteine at a position corresponding to amino acid position 243, leucine at a position corresponding to amino acid position 365 and methionine at a position corresponding to amino acid position 381.

29. An isolated monoclonal antibody specific for and immunoreactive with the isolated polypeptide of claim 28.

30. A method for screening for drugs targeting the isolated polypeptide of claim 28 which comprises contacting the EDG6 polymorphic variant with a candidate agent and assaying for binding activity.

31. A computer system for storing and analyzing polymorphism data for the endothelial differentiation, G-protein-coupled receptor 6 gene, comprising:

(a) a central processing unit (CPU);

(b) a communication interface;

(c) a display device;

(d) an input device; and

(e) a database containing the polymorphism data;

wherein the polymorphism data comprises the haplotypes set forth in the table immediately below:

Haplotype Number[a]										PS[b]	PS
1	2	3	4	5	6	7	8	9	10	Number	Position[c]
A	A	G	G	G	G	G	G	G	G	1	3591
C	C	C	C	C	C	C	C	C	C	2	3697
C	C	C	C	C	C	C	C	C	C	3	3804
G	G	A	A	A	A	A	G	G	G	4	3818
C	C	C	C	C	C	C	C	C	C	5	4123
G	G	A	G	G	G	G	A	G	G	6	4240
G	G	G	G	G	G	G	G	A	G	7	4472
G	G	G	G	G	G	G	G	G	A	8	4499
G	G	G	G	G	G	G	G	G	G	9	4531
G	G	G	G	G	G	G	G	G	G	10	4574
C	C	C	C	C	C	T	C	C	C	11	4736
C	C	C	C	C	C	C	C	C	C	12	4813
C	C	C	C	C	C	C	C	C	C	13	5068
G	G	G	G	G	G	G	G	G	G	14	5103
G	G	G	G	G	G	G	G	G	G	15	5150
G	G	G	G	G	G	G	G	G	G	16	5179
G	G	G	A	G	G	G	G	G	G	17	5301
G	G	G	G	G	G	G	G	G	G	18	5333
G	G	G	G	G	G	G	G	G	G	19	5448
G	G	G	G	G	G	G	G	G	G	20	5560
G	G	G	G	G	G	G	G	G	G	21	5580
C	T	T	C	C	T	T	C	T	C	22	5587
G	C	G	G	G	G	G	G	G	G	23	5606

Haplotype Number[a]										PS[b] Number	PS Position[c]
11	12	13	14	15	16	17	18	19	20		
G	G	G	G	G	G	G	G	G	G	1	3591
C	C	C	C	C	C	C	C	C	C	2	3697
C	C	C	C	C	C	C	C	C	C	3	3804
G	G	G	G	G	G	G	G	G	G	4	3818
C	C	C	C	C	C	C	C	C	C	5	4123
G	G	G	G	G	G	G	G	G	G	6	4240
G	G	G	G	G	G	G	G	G	G	7	4472
G	G	G	G	G	G	G	G	G	G	8	4499
A	G	G	G	G	G	G	G	G	G	9	4531
G	G	G	G	G	G	G	G	G	G	10	4574
C	C	C	C	C	C	C	C	C	C	11	4736
T	C	C	C	C	C	C	C	C	C	12	4813
C	C	C	C	C	C	C	C	T	T	13	5068
G	G	G	G	G	G	G	G	G	G	14	5103
G	A	G	G	G	G	G	G	G	G	15	5150
G	G	G	G	G	G	G	G	A	G	16	5179
G	G	G	G	G	G	G	G	G	G	17	5301
G	G	A	G	G	G	G	G	G	G	18	5333
G	G	G	C	G	G	G	G	G	G	19	5448
G	G	G	G	A	G	G	G	G	G	20	5560
G	G	G	G	G	A	G	G	G	G	21	5580
T	C	C	C	C	C	C	T	C	C	22	5587
G	G	G	G	G	G	G	G	G	G	23	5606

Haplotype Number[a]				PS[b] Number	PS Position[c]
21	22	23	24		
G	G	G	G	1	3591
C	C	C	T	2	3697
C	C	T	C	3	3804
G	G	G	A	4	3818
C	T	C	C	5	4123
G	G	G	G	6	4240
G	G	G	G	7	4472
G	G	G	G	8	4499
G	G	G	G	9	4531
T	G	G	G	10	4574
C	C	C	C	11	4736
C	C	C	C	12	4813
C	C	C	C	13	5068
G	G	G	T	14	5103
G	G	G	G	15	5150
G	G	G	G	16	5179
G	G	G	G	17	5301
G	G	G	G	18	5333
G	G	G	G	19	5448
G	G	G	G	20	5560
G	G	G	G	21	5580
C	C	T	C	22	5587
G	G	G	G	23	5606

[a]Alleles for haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1;

and the haplotype pairs set forth in the table immediately below:

Haplotype Pair[a]								PS[b] Number	PS Position[c]
18/18	17/17	5/5	16/16	17/24	5/7	17/9	17/20		
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/T	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	A/A	G/G	G/A	A/A	G/G	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/T	13	5068
G/G	G/G	G/G	G/G	G/T	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	A/A	G/G	G/G	G/G	G/G	21	5580
T/T	C/C	C/C	C/C	C/C	C/T	C/T	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b] Number	PS Position[c]
17/22	17/1	17/10	17/6	17/12	17/7	17/13	18/10		
G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	G/G	G/A	G/G	G/A	G/G	G/G	4	3818
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/A	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/T	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/A	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	21	5580
C/C	C/C	C/C	C/T	C/C	C/T	C/C	T/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b]	PS
5/3	17/3	5/6	17/8	18/14	17/14	17/5	17/15	Number	Position[c]
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
A/A	G/A	A/A	G/G	G/G	G/G	G/A	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/A	G/A	G/G	G/A	G/G	G/G	G/G	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	10	4574
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	11	4736
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/C	G/C	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/A	20	5560
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	21	5580
C/T	C/T	C/T	C/C	T/C	C/C	C/C	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]								PS[b] Number	PS Position[c]
17/23	17/11	5/4	18/6	17/18	17/16	18/3	17/21		
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	1	3591
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	2	3697
C/T	C/C	C/C	C/C	C/C	C/C	C/C	C/C	3	3804
G/G	G/G	A/A	G/A	G/G	G/G	G/A	G/G	4	3818
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	5	4123
G/G	G/G	G/G	G/G	G/G	G/G	G/A	G/G	6	4240
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	7	4472
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	8	4499
G/G	G/A	G/G	G/G	G/G	G/G	G/G	G/G	9	4531
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/T	10	4574
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	11	4736
C/C	C/T	C/C	C/C	C/C	C/C	C/C	C/C	12	4813
C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C	13	5068
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	14	5103
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	15	5150
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	16	5179
G/G	G/G	G/A	G/G	G/G	G/G	G/G	G/G	17	5301
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	18	5333
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	19	5448
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	20	5560
G/G	G/G	G/G	G/G	G/G	G/A	G/G	G/G	21	5580
C/T	C/T	C/C	T/T	C/T	C/C	T/T	C/C	22	5587
G/G	G/G	G/G	G/G	G/G	G/G	G/G	G/G	23	5606

Haplotype Pair[a]			PS[b] Number	PS Position[c]
18/5	17/2	17/19		
G/G	G/A	G/G	1	3591
C/C	C/C	C/C	2	3697
C/C	C/C	C/C	3	3804
G/A	G/G	G/G	4	3818
C/C	C/C	C/C	5	4123
G/G	G/G	G/G	6	4240
G/G	G/G	G/G	7	4472
G/G	G/G	G/G	8	4499
G/G	G/G	G/G	9	4531
G/G	G/G	G/G	10	4574
C/C	C/C	C/C	11	4736
C/C	C/C	C/C	12	4813
C/C	C/C	C/T	13	5068
G/G	G/G	G/G	14	5103
G/G	G/G	G/G	15	5150
G/G	G/G	G/A	16	5179
G/G	G/G	G/G	17	5301
G/G	G/G	G/G	18	5333
G/G	G/G	G/G	19	5448
G/G	G/G	G/G	20	5560
G/G	G/G	G/G	21	5580
T/C	C/T	C/C	22	5587
G/G	G/C	G/G	23	5606

[a]Haplotype pairs are represented as 1st Haplotype/2nd Haplotype; with alleles of each haplotype shown 5′ to 3′ as 1st polymorphism/2nd polymorphism in each column;

[b]PS = polymorphic site

[c]Location of PS in SEQ ID NO:1.

32. A genome anthology for the endothelial differentiation, G-protein-coupled receptor 6 (EDG6) gene which comprises EDG6 isogenes defined by any one of haplotypes 1-24 set forth in the table shown below:

Haplotype Number[a]										PS[b] Number	PS Position[c]
1	2	3	4	5	6	7	8	9	10		
A	A	G	G	G	G	G	G	G	G	1	3591
C	C	C	C	C	C	C	C	C	C	2	3697
C	C	C	C	C	C	C	C	C	C	3	3804
G	G	A	A	A	A	A	G	G	G	4	3818
C	C	C	C	C	C	C	C	C	C	5	4123
G	G	A	G	G	G	G	A	G	G	6	4240
G	G	G	G	G	G	G	G	A	G	7	4472
G	G	G	G	G	G	G	G	G	A	8	4499
G	G	G	G	G	G	G	G	G	G	9	4531
G	G	G	G	G	G	G	G	G	G	10	4574
C	C	C	C	C	C	T	C	C	C	11	4736
C	C	C	C	C	C	C	C	C	C	12	4813
C	C	C	C	C	C	C	C	C	C	13	5068
G	G	G	G	G	G	G	G	G	G	14	5103
G	G	G	G	G	G	G	G	G	G	15	5150
G	G	G	G	G	G	G	G	G	G	16	5179
G	G	G	A	G	G	G	G	G	G	17	5301
G	G	G	G	G	G	G	G	G	G	18	5333
G	G	G	G	G	G	G	G	G	G	19	5448
G	G	G	G	G	G	G	G	G	G	20	5560
G	G	G	G	G	G	G	G	G	G	21	5580
C	T	T	C	C	T	T	C	T	C	22	5587
G	C	G	G	G	G	G	G	G	G	23	5606

Haplotype Number[a]										PS[b] Number	PS Position[c]
11	12	13	14	15	16	17	18	19	20		
G	G	G	G	G	G	G	G	G	G	1	3591
C	C	C	C	C	C	C	C	C	C	2	3697
C	C	C	C	C	C	C	C	C	C	3	3804
G	G	G	G	G	G	G	G	G	G	4	3818
C	C	C	C	C	C	C	C	C	C	5	4123
G	G	G	G	G	G	G	G	G	G	6	4240
G	G	G	G	G	G	G	G	G	G	7	4472
G	G	G	G	G	G	G	G	G	G	8	4499
A	G	G	G	G	G	G	G	G	G	9	4531
G	G	G	G	G	G	G	G	G	G	10	4574
C	C	C	C	C	C	C	C	C	C	11	4736
T	C	C	C	C	C	C	C	C	C	12	4813
C	C	C	C	C	C	C	C	T	T	13	5068
G	G	G	G	G	G	G	G	G	G	14	5103
G	A	G	G	G	G	G	G	G	G	15	5150
G	G	G	G	G	G	G	G	A	G	16	5179
G	G	G	G	G	G	G	G	G	G	17	5301
G	G	A	G	G	G	G	G	G	G	18	5333
G	G	G	C	G	G	G	G	G	G	19	5448
G	G	G	G	A	G	G	G	G	G	20	5560
G	G	G	G	G	A	G	G	G	G	21	5580
T	C	C	C	C	C	C	T	C	C	22	5587
G	G	G	G	G	G	G	G	G	G	23	5606

Haplotype Number[a]				PS[b]	PS
21	22	23	24	Number	Position[c]
G	G	G	G	1	3591
C	C	C	T	2	3697
C	C	T	C	3	3804
G	G	G	A	4	3818
C	T	C	C	5	4123
G	G	G	G	6	4240
G	G	G	G	7	4472
G	G	G	G	8	4499
G	G	G	G	9	4531
T	G	G	G	10	4574
C	C	C	C	11	4736
C	C	C	C	12	4813
C	C	C	C	13	5068
G	G	G	T	14	5103
G	G	G	G	15	5150
G	G	G	G	16	5179
G	G	G	G	17	5301
G	G	G	G	18	5333
G	G	G	G	19	5448
G	G	G	G	20	5560
G	G	G	G	21	5580
C	C	T	C	22	5587
G	G	G	G	23	5606

[a]Alleles for haplotypes are presented 5′ to 3′ in each column
[b]PS = polymorphic site;
[c]Position of PS in SEQ ID NO:1.

POLYMORPHISMS IN THE EDG6 GENE

ACGCCTGTGT TCTCATGGGA CGCCTGTGTT CTCATGGGAC GCCTGTGCCC
CTCATGGGAC GCCTGTGTTC TCATGGGACG CCTGTGCCCT CATGGGACGC 100
CTGTGTTCTC ATGGGACGCC TGTGCCCCTC ATGGGACGCC TGTGTTCTCA
TGGGACGCCT GTGCCCCTCA TGGGACACCT GTGCCCTCAT GGGACGCCTG 200
TGCCCTCATG GGACGCCTGT GTTCTCATGG GACGCCTGTG CCCCTCATGG
GACGCCTGTG CCCTCATGGG ACGCCTGTGT TCTCATGGGA CGCCTGTGCC 300
CCTCATGGGA CACCTGTGTT CTCATGGGAC GCCTGTGCCC TCATGGGACG
CCTGTGTTCT CATGGGACGC CTGTGCCCCT CATGGGACAC CTGTGTTCTC 400
ATGGCCCCTC ATGGGACACC TGTGTTCTCA TGGGACGCCT GTGCCCTCAT
GGGACGCCTG TGTTCTCATG GGACGCCTGT GCCCCTCATG GGACACCTGT 500
GCCCTCATGG GACGCCTGTG CCCTCATGGG ACGCCTGTGT TCTCATGGGA
CGCCTGTGCC CCTCATGGGA CGCCTGTGTT CTCATGGGAC GCCTGTGCCC 600
CTCATGGGAC ACCTGTGTTC TCATGGGACG CCTGTGCCCT CATGGGACGC
CTGCGTTCTC ATGGGACGCC TGCATTCTCA TGGGACGCCT GTGCCCTCAT 700
GGGACACCTG TGTTCTCATG GGACACCTGT GTTCTCATGG GATGCCTGTG
CCCTCATGGG ATGCCTGTAC CCCTCATGGG ACGCCTGTGT TCTCATGGGA 800
TGCCTGTACC CTCATGGGAC GCCTGTACCC CTCATGGGAC ATCTGTGCTC
TCATGGGATG CCTGTGCCCC TCATGGGATG CCTGTGCCCT CATGGGACGC 900
CTGCATTCTC ATGGGACACC TGTGCCCTCA TGGGATGCCT GTACCCTCAT
GGGATGCCTG TACCCTCATG GGACGCCTGT GCCCCTCATG GGATGCCTGT 1000
GTTCTCATGG GATGCCTGTG CCCCTCATGG GACGCCTGCA TTCTCATGGG
ACACCTGTGC CCTCATGGGA TGCCTGTACC CTTCATGGGA CGCCTGTGTG 1100
TGGTTGCCAT GATTACTACC TGAGACTGTC ACTACGACAG TTACTATTGT
TACTACTTGA GACCATCATT ACAAGACTGA ACGAAGGGAC GAATGTAGAA 1200
ATGAAAACTT AAGACAGAAG AAACTGTTTT AAAGGAAGGG ACCAGGGGAA
GAAAAAGAGA GCTCCCTGCT TCTAGTGAGC AAAGGCAGCC CCCCAAGCTT 1300
CTACAGCCCT TCGTACTTAT TGGGTAGAAA GCAGGGGGAG GAAACGATTG
GCCAGCTGCT TGATTGTTCA CACGTTCACG TTATTGCTAA CAGGTTTCAG 1400
ATTTGCCTAC TTGCAAGAAA CACTTGTGCC TGGGGCGTGA CTGCCCTCAG
CATTCCTTCT GGGCGGCAGA CGCAGTTTGT CAGTTTGCCA ACAGCCTGCT 1500
TTCATGAGAA CAGTTTGCTG TTTACTCACG TAGCCTCCAG TGGTATACTG
AGTTGATCAC AACCCTCATT CTTTCGGCCT TCAACACCTG AGCCCTCACG 1600
GGACATCTGT GCCCCTCATG GGACACCTGT GTCCTCGCAG TACACTTGTG
ACCCTTCCAG GACACCTTAC TGGTAGAATT AGTGTAGCTG CCCCCACCCT 1700
GAGGCCAAGG ACACCATTGT CTCAGGAAGG CTGAAGACCA CAGGCTCCTG
GGGGGACAGA GGGCAGGTGG GGCCCCTCAG GACCCTCCTT GGTGGTAAGT 1800
GGGCCTGGCC TGGGGGTGAT TGCAGGCGGG AGGAGGCTCC CAGCAGGGAC
TTATCCTGGG TCCTACTCAC ACTTCTGGGG CCTGCATTAT TTCCCAAATC 1900
ACCCCACACC CCAAGGCCTT CTGGATGGGG ACGAGTGGGG GGTCACAGAC
ACTGGGGGAG CTGGAGAGCA GAGACCTCAC ACTCCATCCG TGACAGATGA 2000
TGTCCAAGCC CCTACATGCC CCAGACCCCA GGGCAAGGCT GAGCCTCCCT
CCTCAGACCC CAGGGCAAGG CTGAGCATCC CCACTCAGAC CTCAGGGTAG 2100
GGCCTCGCCT CCTCCCTTGG ACCCCAGGAC AGGGCCTCAC CTCTCCCCTC
AGACCCCAAG GCAGGGCCTC GCCTCCCCCC TCAGACTCAG GACAGGGCCA 2200
TGCCTCCCCA CTCAGACCCC AAGGCAAGGC CAACCCTCAC CCCTAGACCC
CAGGCAGGTT CAAGCCTCCC CGCTCAAACC TCAGGGCAGG GCATACCTCC 2300
CTCCTCAGAC CCAGGGCAGG GTGTGTCTCC CCTCTCAGAC CACAGCACAG
GGCCTCGCAT CCCTCCTCAG ACCCCAGGAC AAGGCTGAGC CTCTCCGCTC 2400
AGACCCCAGG GTAAGGTCAT GCCTCCCCTC TCAGAACCTA GGGCAAGGCC
AACCCTCCCC CCTCAGACCC CAGGCAGGTC CAAGCTTCCC CCTCAGACGC 2500
CAGGGCAGAG CCTGCCTCCC TCCTCAGACG CCAGAGCAGG GTGTGTCTCC
CACTTAGACC CACAGCCACC TCACCTCAGG CTGAGTCACC GTGAGCCGTT 2600

FIGURE 1A

```
GTCAGCAGGG CCATGGGGAT GGGGTGAGCA AGTCCCTACT TTTCCTATGC
ACCTCAATCC CAGTGGGGGG CTGCCCCAGG GGGCCAGCAG CTCTGCTCCC      2700
AGCAGGGTGA GCTCAGGGGC AGGAGAGGCA ACCATGAATC CCAAAATGGG
CGCCTGGCCC CGAGACCCTA CCAGCTTGTC CCTGGGGGTC TCTCTCCCTG      2800
GACCTGTGTG TCCTCCCATG GGCAGAAAGT TGGCCTCAGG CCGCCTTTGT
GGGCCTCAGT TTTCCCTTCT GTAACGCATC AGGCTCTCTG GGCCTCAGTT      2900
TCCCCTTCTG CAACGCATCG GGCTCTCTGG GCCTCAGTTT CCCCTTCTGC
AACGCATCGG GCTCTCTGGG CCTCAGTTTC CCCTTCTGCA ACGCATCGGG      3000
CTCTCTGGGC CTCAGTTTCC CCTTCTGCAA CGCCTCGGGC TCTCTGGGCC
TCAGTTTCCC CTTCTGCAAC GCATCGGGCT CTCTGGGCCT CAGTTTCCCC      3100
TTCTGCAACG CCTCGGGCTC TCTGGGCCTC AGTTTCCCCT TCTGCAACGC
CTCGGGCTCT CTGGGCCTCA GTTTCCCCTT CTGCAACGCA TCGGGCTCTC      3200
TGGGCCTCAG TTTCCCCTTC TGCAACGCAT CGGGCTCTCT GGGCCTCAGT
TTCCCCTTCT GCAACGCATC GGGCTCTCTG GCCTCAGTT TCCCCTTCTG      3300
CAACGCCTCG GGCTCTCTGG GCCTCAGTTT CCCCTTCTGC AACGCCTCGG
GCTCTCTGGG CCTCAGTTTC CCCTTCTGCA ACGCATCGGG CTCTCTGGGC      3400
CTCAGTTTCC CCTTCTGCAA CGCATCGGGC TCTCTGGGCC TCAGTTTCCC
CTTCTGCAAC GCACCGGGCT CTCTGGGCCT CAGTTTCCTC TTCTGTAACG      3500
CATCAGGCCC TGTTTCGGGG ATCAAGTCGG ATGAGTCAGT GCTCAAGGGC
ATGCAGGCAC TTGACATTTA TTAGGCACCT GCTGTGTGCT GAGCGCCGGT      3600
                                                    A
GGAGCATGTG GGGAGACCTC AGTGGAGCCG ACGGGTGCTT CGGGGGTGAT
CAGGGCTGGT GGGGAGGAGT CGTCCGGCTG GAAAGGGGTG GCCCATCCCG      3700
                                                      T
AGTGGGGACT CATTTCCCCT CCGTGACTGA CGGCTCCGGG GCTCCCTGCG
GGTCTCCGGC TTTCGAGGAC AGGAAGAAGG CAGCCAGGGC AGGGGTGGGG      3800
GTCCTCACAG CCAGGGCAGC CCCAGCGCGT TGGCTCCAGG AGCCCGGGTG
  T             G
GGGGCCGACC GTTGGGGTGC CCCTCCCTGT CCTCGGCCTT ACCTCCACCC      3900
TGGAGGGCAC CTTGAACATA ACAGGAAATT TCAAATAACA GGAAACCAAG
ACCAGCAAGG CGGGTGGCTC CACCCTGCGT CGGGCCTGAG TCAGCCCCCG      4000
GGGGAGGCCA TGAACGCCAC GGGGACCCCG GTGGCCCCCG AGTCCTGCCA
     [exon 1: 4010..
ACAGCTGGCG GCCGGCGGGC ACAGCCGGCT CATTGTTCTG CACTACAACC      4100
ACTCGGGCCG GCTGGCCGGG CGCGGGGGGC CGGAGGATGG CGGCCTGGGG
                        T
GCCCTGCGGG GGCTGTCGGT GGCCGCCAGC TGCCTGGTGG TGCTGGAGAA      4200
CTTGCTGGTG CTGGCGGCCA TCACCAGCCA CATGCGGTCG CGACGCTGGG
                                          A
TCTACTATTG CCTGGTGAAC ATCACGCTGA GTGACCTGCT CACGGGCGCG      4300
GCCTACCTGG CCAACGTGCT GCTGTCGGGG GCCCGCACCT TCCGTCTGGC
GCCCGCCCAG TGGTTCCTAC GGGAGGGCCT GCTCTTCACC GCCCTGGCCG      4400
CCTCCACCTT CAGCCTGCTC TTCACTGCAG GGGAGCGCTT TGCCACCATG
GTGCGGCCGG TGGCCGAGAG CGGGGCCACC AAGACCAGCC GCGTCTACGG      4500
                  A                                   A
CTTCATCGGC CTCTGCTGGC TGCTGGCCGC GCTGCTGGGG ATGCTGCCTT
                                    A
TGCTGGGCTG GAACTGCCTG TGCGCCTTTG ACCGCTGCTC CAGCCTTCTG      4600
                        T
CCCCTCTACT CCAAGCGCTA CATCCTCTTC TGCCTGGTGA TCTTCGCCGG
CGTCCTGGCC ACCATCATGG GCCTCTATGG GGCCATCTTC CGCCTGGTGC      4700
AGGCCAGCGG GCAGAAGGCC CCACGCCCAG CGGCCCGCCG CAAGGCCCGC
                                          T
CGCCTGCTGA AGACGGTGCT GATGATCCTG CTGGCCTTCC TGGTGTGCTG      4800
```

FIGURE 1B

```
GGGCCCACTC TTCGGGCTGC TGCTGGCCGA CGTCTTTGGC TCCAACCTCT
             T
GGGCCCAGGA GTACCTGCGG GGCATGGACT GGATCCTGGC CCTGGCCGTC      4900
CTCAACTCGG CGGTCAACCC CATCATCTAC TCCTTCCGCA GCAGGGAGGT
GTGCAGAGCC GTGCTCAGCT TCCTCTGCTG CGGGTGTCTC CGGCTGGGCA      5000
TGCGAGGGCC CGGGGACTGC CTGGCCCGGG CCGTCGAGGC TCACTCCGGA
GCTTCCACCA CCGACAGCTC TCTGAGGCCA AGGGACAGCT TTCGCGGCTC      5100
                  T
CCGCTCGCTC AGCTTTCGGA TGCGGGAGCC CCTGTCCAGC ATCTCCAGCG
  T                                                    A
TGCGGAGCAT CTGAAGTTGC AGTCTTGCGT GTGGATGGTG CAGCCACCGG      5200
                           A
                ..5164]
GTGCGTGCCA GGCAGGCCCT CCTGGGGTAC AGGAAGCTGT GTGCACGCAG
CCTCGCCTGT ATGGGGAGCA GGGAACGGGA CAGGCCCCCA TGGTCTTCCC      5300
GGTGGCCTCT CGGGGCTTCT GACGCCAAAT GGGCTTCCCA TGGTCACCCT
A                                  A
GGACAAGGAG GTAACCACCC CACCTCCCCG TAGGAGCAGA GAGCACCCTG      5400
GTGTGGGGGC GAGTGGTTCC CCACAACCCC GCTTCTGTGT GATTCTGGGG
                                                       C
AAGTCCCGGC CCCTCTCTGG GCCTCAGTAG GGCTCCCAGG CTGCAAGGGG      5500
TGGACTGTGG GATGCATGCC CTGGCAACAT TGAAGTTCGA TCATGGTACG
TGATGTTGCG GCCTCTTATT CCCTGGTGCG TGCATGCGTG GGGGCCGTGG      5600
         A                         A       T
CTCAGGGGGG CTGTGGATCT AGGGGCAGCC GGGTGTGTCT TTGCTAGAGA
     C
GGGCCACGGG CCAGTGCCCT GTGAGGGTGG AGTGTGTGTG TGTGTGTGTG      5700
TGTGTGTGTG TGTGTGTGGA CAACCTCTGG GCGTTGCGGG AAGTGGGGGT
GACAATGACA GTTAATGCCG CCTCTTCTTG TTCACTTCCC CTTTAGAAAT      5800
GGCAGGGCCC ATGCCCCATC TCTGGCCTCT GCATCTTTTG GGGACCCACT
CTCTGGGGCT GGCAGAGGCA CCACCTTGGC TTCCTGGGCT GGGGGAATCT      5900
TCCCTCACAT CCCCTTCAGC ATGAACGGCC TCGGCTTTCC CGGTGGGTAA
AACAGTTTAA TCACTGAAGC CGAAGCACAG GGTTGATTGT ACACGCTCCC      6000
CGCCAGCCAC AGGGGCTGAC AACTGCCTGC CCCGTGAAAC TCCAGTGGAG
ACGTTTCAGC TCCACACCAT TCAGTATGGG AGACGCCAGC CCCACGGGGC      6100
TACGGTGCAA GCAGATAACT GAATTTCGAA GTGTAGGTTG TGTTTAATTT
GAATCTGTTT ATATTTCGGT AGCCCCATGG GGCGGGTGGC CACAGTTTCA      6200
GTGCAGATGT AAATCCGGAA GCCTCCAGCA CCTGCAGCTC ATAGACAGCT
CTCGCCCACC TTCTCCCAGG ACCAAGCCAG TCCCGTCCAG TCCAGTGTCT      6300
GAGCAGAGTC AGAATCCACA CCACCCGCCG CCTGGGCTCA GAAAGTTCTG
CTTTAAGTCA TTATTTCTCC ACTGTACGAT GGGGAATGCG GTGTGTGGGG      6400
GCCATTTACC CACCAACACA GCAGCTGTGA GGCACACACG GCTATTGAAA
ATTCATGGAA ATTGCTGGGT GTGGTGGCTC ATGCCTGTAA TCCCAGCACT      6500
TTGGGAGGCC GAGGCAGGAG GATTGCTTGA GTGCAGGAGT TCCAGACCAG
CCTGGGCAAC ATAGCGAAAC CACATCTCTA CAAAAAAATC CTCCAAAATT      6600
AAAAAAATTA GCCCGAGCAT GTTGTTGCAT GCCTGTGGTT CCAGCTACTC
GAGAGGCTGA GGTGGGAGGA TCACTTGAGC CCGGGAGGTC GAGGCTGCAG      6700
GGAGCGGTGA TCGTGCCAGC CTGGGTGACA GAGTGAGACC CTGTCTCTAA
AAAAAAATTA AAAAAAAAAA AAAAAGAAAA TTCATTGAAA TTAAATGAAA      6800
GTTAGGCCAG GTGCATTGGT TCATGCCTGT AATCGCAGCA CTTTGGGGGG
CCGAGGTGGG TGGATCATGA GGTCAGGAGT TCAAGACCAG CCTGGCTAAG      6900
ATGGTGAAAC CCTGTCTCTA CTAAAAATAC AAAAAATAAG CTGGGTGCGG
TGGCAGGCGC CTGTAATCCC AGCTACTCGG GAGGCTGAGG CAGGAGAATC      7000
GCTTGAACTC GGAGGGTGGA GGTTGCAGTG AGCTGGGATT GCTCCACTGC
```

FIGURE 1C

ACTCCAGTCT GGGCGATAGA GTGAGACTCC GTCTCAAAAA ACAAAACAAA 7100
AACAACAACA AAAAAGAAAT TAAATGAAAG TTAAAAATTT AGTCTCTCAG
TGTATTGATT TGCTAGGGCA GCCATGACAG TCTCACAGAT GGAGAGGCTT 7200
GAAAACAGAC ATTTATGCTG CCATAATTCT GGAAGCCAGA AGTCTGAGAT
GAAGGTGTGA GTAGGGCTGG GTCCTCCTGA GGCTGACGGG GATCTGCCCA 7300
GGCCTCTCCC AGTTCCTGCT GGTGGCCGGC GTCCTTGATG TTCCTTGGCT
TGGAGAAGCG TCAGCCGCAT CTCTGCCTCC ATCTCCACAA GGCGCCCTCC 7400
CTGGGTCTGT ATCCAAACTC CCCCAACTCC TTCAACCTTT TTTATTTTAA
GAGATGAGGG GCCGGGCGCC GTGGCTCACG CCTGTAATCC CAGCACTTTG 7500
GGAGGCCGAG GTGGGCGGAT CACGAGGTCA GGAGATAGAG ACCATCCTGG
CTAACACGGT GAAGCCCCAT CTTTACTAAA AATACAAAGA ATTAGCCAGG 7600
CGTGGTGGCG GGTGCCTGTA GTCCCAGCTA CTCTGGAGGC TGAGTCAGGA
GAATGGCGTG AACCCGGGAG GCAGAACTTG CAGTGAGCGG AGATCAGGCC 7700
ACTGCACTCC ACCCTGGGTG ACAGAGAGAG ACTCCGTCTC AAAAAAAAAA
AAAAAAAAAA AAAAAAAAGA GATGAGGGTC TCATTAAGTT GCCCAGGCTG 7800
GTCTCAAACT CCTGGGCTCA AGTGATCCTC CCACCTCAGC CTTCTGAATA
GCTGAGACTA CAGGAGTGCA CCACCAAGCC TGGCTCATTT TCCTATTTTT 7900
ACATTTATTA TTATTATTGT TATTATTATT TTTTTGAGAA GGACTCTCGC
TCTGTCGTCC AGGCTGGAGC GCAGTGGCGC GATCTCAGCT CACCACAACC 8000
TCCACCTCCT GGGTTCAAGC GATTCTCCTG CCTCAGCCTC CCGACTGGGA
CTACAGGCAC GCACCATCAT ATCTGGCTAG TTTTTGTATT TTTAGTAGAG 8100
ACGGGATTTT GCCATGTTGG CCAGGCTGGT CTCGAACTCC TGACTCAGGT
GATCCATCTG CCTCCCAAAG TGCTGGGATT ACAGGCGTGA GCCACCGTGC 8200
CCGGCCGATT TCCCCATTTT TATAAGGCCA CCAGTCACAT GGGATTAAGG
GCCCACCCTG CTCCAGTATG ACCTCATCTT AACTAATGCC ATCTGCCACA 8300
ACCATTTTCC AAATAACATC CTATGCTGAG GTCCTGGGGG TTAGGATGTG
ATCATGTAGA TTTTGATGGG ACAGTTCTTT ATTTTCTTTT ATATTTTAGA 8400
GACAGGGTCT TGCTCTGTCA CCCAGGCTGG AGTGCAGTGG TGTGATCATA
GCTCACTGCA GCCTCGATCT CTTGGGCTCA AGTGATCCTT CTGCCTTGGG 8500
TTTCCGAGTA GCTGCAATTA CAGGTGCACA CCACCACACC TGGCTAATTT
TTATATTTTT TGTAGAGACG GGGCTGGGTG CTGTGGCTCA CCCCTGTAAT 8600
TCCAGCACTT TGGGAGGCCG AGGCAGGTGG ATCACCTTAG GTCAGGAGTT
TGAGACCAGC CTGGCCAACA TGGAGAAACC CCGTCTCTAC TAAAAATACA 8700
AAATTAGCCA GGCGTAGTGG TGCATGCCTG TAATCCCAGC TACTTGGGAG
GCTGAGGCAG GAGAATTGCT TGAACCCGGG AGGCCGAGGT TGCAGTGGAC 8800
TGAGTTCGTG CCACTGCACT CCAATCTGGG TGACAGAGCA AGACTCCGTC
TCAAAAAAAA AAAAAAAAAA AAAAGAGAGA GAGAGAGAGA GACAGGGTCT 8900
TGCTATGTTT CCCAGGCTGG TCTCAATCTC CTGGGCTCAA GTGATCCTCC
TGCCTCAGCC TCCCAAAGTG TTGGGACTAC AGGCATGAGC TACTGTGGCT 9000
GGCCAATTGC AGACAGATGG AACTGTGGCC AGCCACTTCA CCAGTGAAGT
CCCAAACACC CCTCTGCAGG CATGATAAGC TAGGGTAGTT CTTATCTTTT 9100
CCTGATTCCT GTTCTCTGTG TGTTTAGTGG GGACCACAGC AGGCAGGGTG
GGGTGTTGGT GGA 9163

FIGURE 1D

POLYMORPHISMS IN THE CODING SEQUENCE OF EDG6

```
ATGAACGCCA CGGGGACCCC GGTGGCCCCC GAGTCCTGCC AACAGCTGGC
GGCCGGCGGG CACAGCCGGC TCATTGTTCT GCACTACAAC CACTCGGGCC      100
GGCTGGCCGG GCGCGGGGGG CCGGAGGATG GCGGCCTGGG GGCCCTGCGG
             T
GGGCTGTCGG TGGCCGCCAG CTGCCTGGTG GTGCTGGAGA ACTTGCTGGT      200
GCTGGCGGCC ATCACCAGCC ACATGCGGTC GCGACGCTGG GTCTACTATT
                                   A
GCCTGGTGAA CATCACGCTG AGTGACCTGC TCACGGGCGC GGCCTACCTG      300
GCCAACGTGC TGCTGTCGGG GGCCCGCACC TTCCGTCTGG CGCCCGCCCA
GTGGTTCCTA CGGGAGGGCC TGCTCTTCAC CGCCCTGGCC GCCTCCACCT      400
TCAGCCTGCT CTTCACTGCA GGGGAGCGCT TTGCCACCAT GGTGCGGCCG
GTGGCCGAGA GCGGGGCCAC CAAGACCAGC CGCGTCTACG GCTTCATCGG      500
             A                                 A
CCTCTGCTGG CTGCTGGCCG CGCTGCTGGG GATGCTGCCT TTGCTGGGCT
                       A
GGAACTGCCT GTGCGCCTTT GACCGCTGCT CCAGCCTTCT GCCCCTCTAC      600
               T
TCCAAGCGCT ACATCCTCTT CTGCCTGGTG ATCTTCGCCG GCGTCCTGGC
CACCATCATG GGCCTCTATG GGGCCATCTT CCGCCTGGTG CAGGCCAGCG      700
GGCAGAAGGC CCCACGCCCA GCGGCCCGCC GCAAGGCCCG CCGCCTGCTG
                             T
AAGACGGTGC TGATGATCCT GCTGGCCTTC CTGGTGTGCT GGGGCCCACT      800
CTTCGGGCTG CTGCTGGCCG ACGTCTTTGG CTCCAACCTC TGGGCCCAGG
   T
AGTACCTGCG GGGCATGGAC TGGATCCTGG CCCTGGCCGT CCTCAACTCG      900
GCGGTCAACC CCATCATCTA CTCCTTCCGC AGCAGGGAGG TGTGCAGAGC
CGTGCTCAGC TTCCTCTGCT GCGGGTGTCT CCGGCTGGGC ATGCGAGGGC     1000
CCGGGGACTG CCTGGCCCGG GCCGTCGAGG CTCACTCCGG AGCTTCCACC
ACCGACAGCT CTCTGAGGCC AAGGGACAGC TTTCGCGGCT CCCGCTCGCT     1100
       T                                      T
CAGCTTTCGG ATGCGGGAGC CCCTGTCCAG CATCTCCAGC GTGCGGAGCA
                                            A
TCTGA                                                      1155
```

FIGURE 2

ISOFORMS OF THE EDG6 PROTEIN

```
MNATGTPVAP ESCQQLAAGG HSRLIVLHYN HSGRLAGRGG PEDGGLGALR
GLSVAASCLV VLENLLVLAA ITSHMRSRRW VYYCLVNITL SDLLTGAAYL      100
ANVLLSGART FRLAPAQWFL REGLLFTALA ASTFSLLFTA GERFATMVRP
VAESGATKTS RVYGFIGLCW LLAALLGMLP LLGWNCLCAF DRCSSLLPLY      200
    R        S                          S
SKRYILFCLV IFAGVLATIM GLYGAIFRLV QASGQKAPRP AARRKARRLL
                                          C
KTVLMILLAF LVCWGPLFGL LLADVFGSNL WAQEYLRGMD WILALAVLNS      300
AVNPIIYSFR SREVCRAVLS FLCCGCLRLG MRGPGDCLAR AVEAHSGAST
TDSSLRPRDS FRGSRSLSFR MREPLSSISS VRSI                       384
              L                  M
```

FIGURE 3

SEQUENCE LISTING

<110> Genaissance Pharmaceuticals Inc.
Kliem, Stefanie E.
Koshy, Beena

<120> Haplotypes of the EDG6 Gene

<130> MWH-0934PCT EDG6

<140> TBA
<141> 2001-07-17

<150> 60/218,727
<151> 2000-07-17

<160> 119

<170> PatentIn Ver. 2.1

<210> 1
<211> 9163
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (3591)
<223> PS1: polymorphic base G or A

<220>
<221> allele
<222> (3697)
<223> PS2: polymorphic base C or T

<220>
<221> allele
<222> (3804)
<223> PS3: polymorphic base C or T

<220>
<221> allele
<222> (3818)
<223> PS4: polymorphic base A or G

<220>
<221> allele
<222> (4123)

<223> PS5: polymorphic base C or T

<220>
<221> allele
<222> (4240)
<223> PS6: polymorphic base G or A

<220>
<221> allele
<222> (4472)
<223> PS7: polymorphic base G or A

<220>
<221> allele
<222> (4499)
<223> PS8: polymorphic base G or A

<220>
<221> allele
<222> (4531)
<223> PS9: polymorphic base, G or A

<220>
<221> allele
<222> (4574)
<223> PS10: polymorphic base G or T

<220>
<221> allele
<222> (4736)
<223> PS11: polymorphic base C or T

<220>
<221> allele
<222> (4813)
<223> PS12: polymorphic base C or T

<220>
<221> allele
<222> (5068)
<223> PS13: polymorphic base C or T

<220>
<221> allele
<222> (5103)
<223> PS14: polymorphic base G or T

<220>

<221> allele
<222> (5150)
<223> PS15: polymorphic base G or A

<220>
<221> allele
<222> (5179)
<223> PS16: polymorphic base G or A

<220>
<221> allele
<222> (5301)
<223> PS17: polymorphic base G or A

<220>
<221> allele
<222> (5333)
<223> PS18: polymorphic base G or A

<220>
<221> allele
<222> (5448)
<223> PS19: polymorphic base G or C

<220>
<221> allele
<222> (5560)
<223> PS20: polymorphic base G or A

<220>
<221> allele
<222> (5580)
<223> PS21 : polymorphic base G or A

<220>
<221> allele
<222> (5587)
<223> PS22 : polymorphic base C or T

<220>
<221> allele
<222> (5606)
<223> PS23 : polymorphic base G or C

<400> 1
acgcctgtgt tctcatggga cgcctgtgtt ctcatgggac gcctgtgccc ctcatgggac 60
gcctgtgttc tcatgggacg cctgtgccct catgggacgc ctgtgttctc atgggacgcc 120
tgtgcccctc atgggacgcc tgtgttctca tgggacgcct gtgcccctca tgggacacct 180

gtgccctcat gggacgcctg tgccctcatg ggacgcctgt gttctcatgg gacgcctgtg 240
cccctcatgg gacgcctgtg ccctcatggg acgcctgtgt tctcatggga cgcctgtgcc 300
cctcatggga cacctgtgtt ctcatgggac gcctgtgccc tcatgggacg cctgtgttct 360
catgggacgc ctgtgcccct catgggacac ctgtgttctc atggcccctc atgggacacc 420
tgtgttctca tgggacgcct gtgccctcat gggacgcctg tgttctcatg ggacgcctgt 480
gcccctcatg ggacacctgt gccctcatgg gacgcctgtg ccctcatggg acgcctgtgt 540
tctcatggga cgcctgtgcc cctcatggga cgcctgtgtt ctcatgggac gcctgtgccc 600
ctcatgggac acctgtgttc tcatgggacg cctgtgccct catgggacgc ctgcgttctc 660
atgggacgcc tgcattctca tgggacgcct gtgccctcat gggacacctg tgttctcatg 720
ggacacctgt gttctcatgg gatgcctgtg ccctcatggg atgcctgtac ccctcatggg 780
acgcctgtgt tctcatggga tgcctgtacc ctcatgggac gcctgtaccc ctcatgggac 840
atctgtgctc tcatgggatg cctgtgcccc tcatgggatg cctgtgccct catgggacgc 900
ctgcattctc atgggacacc tgtgccctca tgggatgcct gtaccctcat gggatgcctg 960
taccctcatg ggacgcctgt gcccctcatg ggatgcctgt gttctcatgg gatgcctgtg 1020
cccctcatgg gacgcctgca ttctcatggg acacctgtgc cctcatggga tgcctgtacc 1080
cttcatggga cgcctgtgtg tggttgccat gattactacc tgagactgtc actacgacag 1140
ttactattgt tactacttga gaccatcatt acaagactga acgaagggac gaatgtagaa 1200
atgaaaactt aagacagaag aaactgtttt aaaggaaggg accaggggaa gaaaaagaga 1260
gctccctgct tctagtgagc aaaggcagcc ccccaagctt ctacagccct tcgtacttat 1320
tgggtagaaa gcaggggggag gaaacgattg gccagctgct tgattgttca cacgttcacg 1380
ttattgctaa caggtttcag atttgcctac ttgcaagaaa cacttgtgcc tggggcgtga 1440
ctgccctcag cattccttct gggcggcaga cgcagtttgt cagtttgcca acagcctgct 1500
ttcatgagaa cagtttgctg tttactcacg tagcctccag tggtatactg agttgatcac 1560
aaccctcatt ctttcggcct tcaacacctg agccctcacg ggacatctgt gcccctcatg 1620
ggacacctgt gtcctcgcag tacacttgtg acccttccag gacaccttac tggtagaatt 1680
agtgtagctg cccccaccct gaggccaagg acaccattgt ctcaggaagg ctgaagacca 1740
caggctcctg gggggacaga gggcaggtgg ggcccctcag gaccctcctt ggtggtaagt 1800
gggcctggcc tgggggtgat tgcaggcggg aggaggctcc cagcagggac ttatcctggg 1860
tcctactcac acttctgggg cctgcattat ttcccaaatc accccacacc ccaaggcctt 1920
ctggatgggg acgagtgggg ggtcacagac actgggggag ctggagagca gagacctcac 1980
actccatccg tgacagatga tgtccaagcc cctacatgcc ccagacccca gggcaaggct 2040
gagcctccct cctcagaccc cagggcaagg ctgagcatcc ccactcagac ctcagggtag 2100
ggcctcgcct cctcccttgg accccaggac agggcctcac ctctcccctc agaccccaag 2160
gcagggcctc gcctcccccc tcagactcag gacagggcca tgcctcccca ctcagacccc 2220
aaggcaaggc caaccctcac ccctagaccc caggcaggtt caagcctccc cgctcaaacc 2280
tcagggcagg gcatacctcc ctcctcagac ccagggcagg gtgtgtctcc cctctcagac 2340
cacagcacag ggcctcgcat ccctcctcag accccaggac aaggctgagc ctctccgctc 2400
agaccccagg gtaaggtcat gcctcccctc tcagaaccta gggcaaggcc aaccctcccc 2460
cctcagaccc caggcaggtc caagcttccc cctcagacgc cagggcagag cctgcctccc 2520
tcctcagacg ccagagcagg gtgtgtctcc cacttagacc cacagccacc tcacctcagg 2580
ctgagtcacc gtgagccgtt gtcagcaggg ccatggggat ggggtgagca agtccctact 2640
tttcctatgc acctcaatcc cagtgggggg ctgccccagg gggccagcag ctctgctccc 2700
agcagggtga gctcaggggc aggagaggca accatgaatc ccaaaatggg cgcctggccc 2760
cgagacccta ccagcttgtc cctgggggtc tctctccctg gacctgtgtg tcctcccatg 2820
ggcagaaagt tggcctcagg ccgcctttgt gggcctcagt tttcccttct gtaacgcatc 2880
aggctctctg ggcctcagtt tccccttctg caacgcatcg ggctctctgg gcctcagttt 2940
cccccttctgc aacgcatcgg gctctctggg cctcagtttc cccttctgca acgcatcggg 3000
ctctctgggc ctcagtttcc ccttctgcaa cgcctcgggc tctctgggcc tcagtttccc 3060

cttctgcaac gcatcgggct ctctgggcct cagtttcccc ttctgcaacg cctcgggctc 3120
tctgggcctc agtttcccct tctgcaacgc ctcgggctct ctgggcctca gtttcccctt 3180
ctgcaacgca tcgggctctc tgggcctcag tttccccttc tgcaacgcat cgggctctct 3240
gggcctcagt ttccccttct gcaacgcatc gggctctctg ggcctcagtt tccccttctg 3300
caacgcctcg ggctctctgg gcctcagttt cccccttctgc aacgcctcgg gctctctggg 3360
cctcagtttc cccttctgca acgcatcggg ctctctgggc ctcagtttcc ccttctgcaa 3420
cgcatcgggc tctctgggcc tcagtttccc cttctgcaac gcaccgggct ctctgggcct 3480
cagtttcctc ttctgtaacg catcaggccc tgtttcgggg atcaagtcgg atgagtcagt 3540
gctcaagggc atgcaggcac ttgacattta ttaggcacct gctgtgtgct ragcgccggt 3600
ggagcatgtg gggagacctc agtggagccg acgggtgctt cgggggtgat cagggctggt 3660
ggggaggagt cgtccggctg gaaaggggtg gcccatyccg agtggggact catttcccct 3720
ccgtgactga cggctccggg gctccctgcg ggtctccggc tttcgaggac aggaagaagg 3780
cagccagggc aggggtgggg gtcytcacag ccagggcrgc cccagcgcgt tggctccagg 3840
agcccgggtg ggggccgacc gttggggtgc ccctccctgt cctcggcctt acctccaccc 3900
tggagggcac cttgaacata acaggaaatt tcaaataaca ggaaaccaag accagcaagg 3960
cgggtggctc caccctgcgt cgggcctgag tcagcccccg ggggaggcca tgaacgccac 4020
ggggaccccg gtggcccccg agtcctgcca acagctggcg gccggcgggc acagccggct 4080
cattgttctg cactacaacc actcgggccg gctggccggg cgyggggggc cggaggatgg 4140
cggcctgggg gccctgcggg ggctgtcggt ggccgccagc tgcctggtgg tgctggagaa 4200
cttgctggtg ctggcggcca tcaccagcca catgcggtcr cgacgctggg tctactattg 4260
cctggtgaac atcacgctga gtgacctgct cacgggcgcg gcctacctgg ccaacgtgct 4320
gctgtcgggg gcccgcacct tccgtctggc gcccgcccag tggttcctac gggagggcct 4380
gctcttcacc gccctggccg cctccacctt cagcctgctc ttcactgcag gggagcgctt 4440
tgccaccatg gtgcggccgg tggccgagag crgggccacc aagaccagcc gcgtctacrg 4500
cttcatcggc ctctgctggc tgctggccgc rctgctgggg atgctgcctt tgctgggctg 4560
gaactgcctg tgckcctttg accgctgctc cagccttctg cccctctact ccaagcgcta 4620
catcctcttc tgcctggtga tcttcgccgg cgtcctggcc accatcatgg gcctctatgg 4680
ggccatcttc cgcctggtgc aggccagcgg gcagaaggcc ccacgcccag cggccygccg 4740
caaggcccgc cgcctgctga agacggtgct gatgatcctg ctggccttcc tggtgtgctg 4800
gggcccactc ttyggggctgc tgctggccga cgtctttggc tccaacctct gggcccagga 4860
gtacctgcgg ggcatggact ggatcctggc cctggccgtc ctcaactcgg cggtcaaccc 4920
catcatctac tccttccgca gcagggaggt gtgcagagcc gtgctcagct tcctctgctg 4980
cgggtgtctc cggctgggca tgcgagggcc cggggactgc ctggcccggg ccgtcgaggc 5040
tcactccgga gcttccacca ccgacagytc tctgaggcca agggacagct ttcgcggctc 5100
cckctcgctc agctttcgga tgcgggagcc cctgtccagc atctccagcr tgcggagcat 5160
ctgaagttgc agtcttgcrt gtggatggtg cagccaccgg gtgcgtgcca ggcaggccct 5220
cctggggtac aggaagctgt gtgcacgcag cctcgcctgt atggggagca gggaacggga 5280
caggccccca tggtcttccc rgtggcctct cggggcttct gacgccaaat ggrcttccca 5340
tggtcaccct ggacaaggag gtaaccaccc cacctccccg taggagcaga gagcaccctg 5400
gtgtgggggc gagtggttcc ccacaacccc gcttctgtgt gattctgsgg aagtcccggc 5460
ccctctctgg gcctcagtag ggctcccagg ctgcaagggg tggactgtgg gatgcatgcc 5520
ctggcaacat tgaagttcga tcatggtacg tgatgttgcr gcctcttatt ccctggtgcr 5580
tgcatgygtg ggggccgtgg ctcagsgggg ctgtggatct aggggcagcc gggtgtgtct 5640
ttgctagaga gggccacggg ccagtgccct gtgagggtgg agtgtgtgtg tgtgtgtgtg 5700
tgtgtgtgtg tgtgtgtgga caacctctgg gcgttgcggg aagtgggggt gacaatgaca 5760
gttaatgccg cctcttcttg ttcacttccc ctttagaaat ggcagggccc atgccccatc 5820
tctggcctct gcatcttttg gggacccact ctctggggct ggcagaggca ccaccttggc 5880
ttcctgggct gggggaatct tccctcacat ccccttcagc atgaacggcc tcggctttcc 5940

cggtgggtaa aacagtttaa tcactgaagc cgaagcacag ggttgattgt acacgctccc 6000
cgccagccac aggggctgac aactgcctgc cccgtgaaac tccagtggag acgtttcagc 6060
tccacaccat tcagtatggg agacgccagc cccacggggc tacggtgcaa gcagataact 6120
gaatttcgaa gtgtaggttg tgtttaattt gaatctgttt atatttcggt agccccatgg 6180
ggcgggtggc cacagtttca gtgcagatgt aaatccggaa gcctccagca cctgcagctc 6240
atagacagct ctcgcccacc ttctcccagg accaagccag tcccgtccag tccagtgtct 6300
gagcagagtc agaatccaca ccacccgccg cctgggctca gaaagttctg ctttaagtca 6360
ttatttctcc actgtacgat ggggaatgcg gtgtgtgggg gccatttacc caccaacaca 6420
gcagctgtga ggcacacacg gctattgaaa attcatggaa attgctgggt gtggtggctc 6480
atgcctgtaa tcccagcact ttgggaggcc gaggcaggag gattgcttga gtgcaggagt 6540
tccagaccag cctgggcaac atagcgaaac cacatctcta caaaaaaatc ctccaaaatt 6600
aaaaaaatta gcccgagcat gttgttgcat gcctgtggtt ccagctactc gagaggctga 6660
ggtgggagga tcacttgagc ccgggaggtc gaggctgcag ggagcggtga tcgtgccagc 6720
ctgggtgaca gagtgagacc ctgtctctaa aaaaaaatta aaaaaaaaaa aaaaagaaaa 6780
ttcattgaaa ttaaatgaaa gttaggccag gtgcattggt tcatgcctgt aatcgcagca 6840
ctttgggggg ccgaggtggg tggatcatga ggtcaggagt tcaagaccag cctggctaag 6900
atggtgaaac cctgtctcta ctaaaaatac aaaaaataag ctgggtgcgg tggcaggcgc 6960
ctgtaatccc agctactcgg gaggctgagg caggagaatc gcttgaactc ggagggtgga 7020
ggttgcagtg agctgggatt gctccactgc actccagtct gggcgataga gtgagactcc 7080
gtctcaaaaa acaaaacaaa aacaacaaca aaaaagaaat taaatgaaag ttaaaaattt 7140
agtctctcag tgtattgatt tgctagggca gccatgacag tctcacagat ggagaggctt 7200
gaaaacagac atttatgctg ccataattct ggaagccaga agtctgagat gaaggtgtga 7260
gtagggctgg gtcctcctga ggctgacggg gatctgccca ggcctctccc agttcctgct 7320
ggtggccggc gtccttgatg ttccttggct tggagaagcg tcagccgcat ctctgcctcc 7380
atctccacaa ggcgccctcc ctgggtctgt atccaaactc ccccaactcc ttcaaccttt 7440
tttattttaa gagatgaggg gccgggcgcc gtggctcacg cctgtaatcc cagcactttg 7500
ggaggccgag gtgggcggat cacgaggtca ggagatagag accatcctgg ctaacacggt 7560
gaagccccat ctttactaaa aatacaaaga attagccagg cgtggtggcg ggtgcctgta 7620
gtcccagcta ctctggaggc tgagtcagga gaatggcgtg aacccgggag gcagaacttg 7680
cagtgagcgg agatcaggcc actgcactcc accctgggtg acagagagag actccgtctc 7740
aaaaaaaaaa aaaaaaaaaa aaaaaaaaga gatgagggtc tcattaagtt gcccaggctg 7800
gtctcaaact cctgggctca agtgatcctc ccacctcagc cttctgaata gctgagacta 7860
caggagtgca ccaccaagcc tggctcattt tcctattttt acatttatta ttattattgt 7920
tattattatt tttttgagaa ggactctcgc tctgtcgtcc aggctggagc gcagtggcgc 7980
gatctcagct caccacaacc tccacctcct gggttcaagc gattctcctg cctcagcctc 8040
ccgactggga ctacaggcac gcaccatcat atctggctag tttttgtatt tttagtagag 8100
acgggatttt gccatgttgg ccaggctggt ctcgaactcc tgactcaggt gatccatctg 8160
cctcccaaag tgctgggatt acaggcgtga gccaccgtgc ccggccgatt tccccatttt 8220
tataaggcca ccagtcacat gggattaagg gcccaccctg ctccagtatg acctcatctt 8280
aactaatgcc atctgccaca accattttcc aaataacatc ctatgctgag gtcctggggg 8340
ttaggatgtg atcatgtaga ttttgatggg acagttcttt attttctttt atattttaga 8400
gacagggtct tgctctgtca cccaggctgg agtgcagtgg tgtgatcata gctcactgca 8460
gcctcgatct cttgggctca agtgatcctt ctgccttggg tttccgagta gctgcaatta 8520
caggtgcaca ccaccacacc tggctaattt ttatattttt tgtagagacg gggctgggtg 8580
ctgtggctca cccctgtaat tccagcactt tgggaggccg aggcaggtgg atcaccttag 8640
gtcaggagtt tgagaccagc ctggccaaca tggagaaacc ccgtctctac taaaaataca 8700
aaattagcca ggcgtagtgg tgcatgcctg taatcccagc tacttgggag gctgaggcag 8760
gagaattgct tgaacccggg aggccgaggt tgcagtggac tgagttcgtg ccactgcact 8820

ccaatctggg tgacagagca agactccgtc tcaaaaaaaa aaaaaaaaaa aaaagagaga 8880
gagagagaga gacagggtct tgctatgttt cccaggctgg tctcaatctc ctgggctcaa 8940
gtgatcctcc tgcctcagcc tcccaaagtg ttgggactac aggcatgagc tactgtggct 9000
ggccaattgc agacagatgg aactgtggcc agccacttca ccagtgaagt cccaaacacc 9060
cctctgcagg catgataagc tagggtagtt cttatctttt cctgattcct gttctctgtg 9120
tgtttagtgg ggaccacagc aggcagggtg gggtgttggt gga 9163

<210> 2
<211> 1155
<212> DNA
<213> Homo sapien

<400> 2
atgaacgcca cggggacccc ggtggccccc gagtcctgcc aacagctggc ggccggcggg 60
cacagccggc tcattgttct gcactacaac cactcgggcc ggctggccgg gcgcgggggg 120
ccggaggatg gcggcctggg ggccctgcgg gggctgtcgg tggccgccag ctgcctggtg 180
gtgctggaga acttgctggt gctggcggcc atcaccagcc acatgcggtc gcgacgctgg 240
gtctactatt gcctggtgaa catcacgctg agtgacctgc tcacgggcgc ggcctacctg 300
gccaacgtgc tgctgtcggg ggcccgcacc ttccgtctgg cgcccgccca gtggttccta 360
cgggagggcc tgctcttcac cgccctggcc gcctccacct tcagcctgct cttcactgca 420
ggggagcgct ttgccaccat ggtgcggccg gtggccgaga gcggggccac caagaccagc 480
cgcgtctacg gcttcatcgg cctctgctgg ctgctggccg cgctgctggg gatgctgcct 540
ttgctgggct ggaactgcct gtgcgccttt gaccgctgct ccagccttct gcccctctac 600
tccaagcgct acatcctctt ctgcctggtg atcttcgccg gcgtcctggc caccatcatg 660
ggcctctatg gggccatctt ccgcctggtg caggccagcg ggcagaaggc cccacgccca 720
gcggcccgcc gcaaggcccg ccgcctgctg aagacggtgc tgatgatcct gctggccttc 780
ctggtgtgct ggggcccact cttcgggctg ctgctggccg acgtctttgg ctccaacctc 840
tgggcccagg agtacctgcg gggcatggac tggatcctgg ccctggccgt cctcaactcg 900
gcggtcaacc ccatcatcta ctccttccgc agcagggagg tgtgcagagc cgtgctcagc 960
ttcctctgct gcgggtgtct ccggctgggc atgcgagggc ccggggactg cctggcccgg 1020
gccgtcgagg ctcactccgg agcttccacc accgacagct ctctgaggcc aagggacagc 1080
tttcgcggct cccgctcgct cagctttcgg atgcgggagc ccctgtccag catctccagc 1140
gtgcggagca tctga 1155

<210> 3
<211> 384
<212> PRT
<213> Homo sapien

<400> 3
Met Asn Ala Thr Gly Thr Pro Val Ala Pro Glu Ser Cys Gln Gln Leu
1 5 10 15

Ala Ala Gly Gly His Ser Arg Leu Ile Val Leu His Tyr Asn His Ser
20 25 30

```
Gly Arg Leu Ala Gly Arg Gly Gly Pro Glu Asp Gly Gly Leu Gly Ala
         35                  40                  45

Leu Arg Gly Leu Ser Val Ala Ala Ser Cys Leu Val Val Leu Glu Asn
     50                  55                  60

Leu Leu Val Leu Ala Ala Ile Thr Ser His Met Arg Ser Arg Arg Trp
 65                  70                  75                  80

Val Tyr Tyr Cys Leu Val Asn Ile Thr Leu Ser Asp Leu Leu Thr Gly
                 85                  90                  95

Ala Ala Tyr Leu Ala Asn Val Leu Leu Ser Gly Ala Arg Thr Phe Arg
            100                 105                 110

Leu Ala Pro Ala Gln Trp Phe Leu Arg Glu Gly Leu Leu Phe Thr Ala
        115                 120                 125

Leu Ala Ala Ser Thr Phe Ser Leu Leu Phe Thr Ala Gly Glu Arg Phe
    130                 135                 140

Ala Thr Met Val Arg Pro Val Ala Glu Ser Gly Ala Thr Lys Thr Ser
145                 150                 155                 160

Arg Val Tyr Gly Phe Ile Gly Leu Cys Trp Leu Leu Ala Ala Leu Leu
                165                 170                 175

Gly Met Leu Pro Leu Leu Gly Trp Asn Cys Leu Cys Ala Phe Asp Arg
            180                 185                 190

Cys Ser Ser Leu Leu Pro Leu Tyr Ser Lys Arg Tyr Ile Leu Phe Cys
        195                 200                 205

Leu Val Ile Phe Ala Gly Val Leu Ala Thr Ile Met Gly Leu Tyr Gly
    210                 215                 220

Ala Ile Phe Arg Leu Val Gln Ala Ser Gly Gln Lys Ala Pro Arg Pro
225                 230                 235                 240

Ala Ala Arg Arg Lys Ala Arg Arg Leu Leu Lys Thr Val Leu Met Ile
                245                 250                 255

Leu Leu Ala Phe Leu Val Cys Trp Gly Pro Leu Phe Gly Leu Leu Leu
            260                 265                 270

Ala Asp Val Phe Gly Ser Asn Leu Trp Ala Gln Glu Tyr Leu Arg Gly
        275                 280                 285
```

```
Met Asp Trp Ile Leu Ala Leu Ala Val Leu Asn Ser Ala Val Asn Pro
    290                 295                 300

Ile Ile Tyr Ser Phe Arg Ser Arg Glu Val Cys Arg Ala Val Leu Ser
305                 310                 315                 320

Phe Leu Cys Cys Gly Cys Leu Arg Leu Gly Met Arg Gly Pro Gly Asp
                325                 330                 335

Cys Leu Ala Arg Ala Val Glu Ala His Ser Gly Ala Ser Thr Thr Asp
            340                 345                 350

Ser Ser Leu Arg Pro Arg Asp Ser Phe Arg Gly Ser Arg Ser Leu Ser
        355                 360                 365

Phe Arg Met Arg Glu Pro Leu Ser Ser Ile Ser Ser Val Arg Ser Ile
    370                 375                 380
```

<210> 4
<211> 15
<212> DNA
<213> Homo sapien

<400> 4
gtgtgctrag cgccg 15

<210> 5
<211> 15
<212> DNA
<213> Homo sapien

<400> 5
ggcccatycc gagtg 15

<210> 6
<211> 15
<212> DNA
<213> Homo sapien

<400> 6
gggggtcytc acagc 15

<210> 7
<211> 15
<212> DNA
<213> Homo sapien

<400> 7
ccagggcrgc cccag 15

<210> 8
<211> 15
<212> DNA
<213> Homo sapien

<400> 8
ccgggcgygg ggggc 15

<210> 9
<211> 15
<212> DNA
<213> Homo sapien

<400> 9
tgcggtcrcg acgct 15

<210> 10
<211> 15
<212> DNA
<213> Homo sapien

<400> 10
cgagagcrgg gccac 15

<210> 11
<211> 15
<212> DNA
<213> Homo sapien

<400> 11
cgtctacrgc ttcat 15

<210> 12
<211> 15

<212> DNA
<213> Homo sapien

<400> 12
tggccgcrct gctgg 15

<210> 13
<211> 15
<212> DNA
<213> Homo sapien

<400> 13
cctgtgckcc tttga 15

<210> 14
<211> 15
<212> DNA
<213> Homo sapien

<400> 14
agcggccygc cgcaa 15

<210> 15
<211> 15
<212> DNA
<213> Homo sapien

<400> 15
cactcttygg gctgc 15

<210> 16
<211> 15
<212> DNA
<213> Homo sapien

<400> 16
ccgacagytc tctga 15

<210> 17
<211> 15
<212> DNA
<213> Homo sapien

<400> 17
ggctccckct cgctc 15

<210> 18
<211> 15
<212> DNA
<213> Homo sapien

<400> 18
ctccagcrtg cggag 15

<210> 19
<211> 15
<212> DNA
<213> Homo sapien

<400> 19
gtcttgcrtg tggat 15

<210> 20
<211> 15
<212> DNA
<213> Homo sapien

<400> 20
tcttcccrgt ggcct 15

<210> 21
<211> 15
<212> DNA
<213> Homo sapien

<400> 21
caaatggrct tccca 15

<210> 22
<211> 15
<212> DNA
<213> Homo sapien

<400> 22
gattctgsgg aagtc 15

<210> 23
<211> 15
<212> DNA
<213> Homo sapien

<400> 23
atgttgcrgc ctctt 15

<210> 24
<211> 15
<212> DNA
<213> Homo sapien

<400> 24
ctggtgcrtg catgc 15

<210> 25
<211> 15
<212> DNA
<213> Homo sapien

<400> 25
gtgcatgygt ggggg 15

<210> 26
<211> 15
<212> DNA
<213> Homo sapien

<400> 26
ggctcagsgg ggctg 15

<210> 27
<211> 15
<212> DNA
<213> Homo sapien

<400> 27
cctgctgtgt gctra 15

<210> 28
<211> 15

<212> DNA
<213> Homo sapien

<400> 28
ctccaccggc gctya 15

<210> 29
<211> 15
<212> DNA
<213> Homo sapien

<400> 29
aggggtggcc catyc 15

<210> 30
<211> 15
<212> DNA
<213> Homo sapien

<400> 30
agtccccact cggra 15

<210> 31
<211> 15
<212> DNA
<213> Homo sapien

<400> 31
aggggtgggg gtcyt 15

<210> 32
<211> 15
<212> DNA
<213> Homo sapien

<400> 32
gccctggctg tgarg 15

<210> 33
<211> 15
<212> DNA
<213> Homo sapien

<400> 33
tcacagccag ggcrg 15

<210> 34
<211> 15
<212> DNA
<213> Homo sapien

<400> 34
aacgcgctgg ggcyg 15

<210> 35
<211> 15
<212> DNA
<213> Homo sapien

<400> 35
ggctggccgg gcgyg 15

<210> 36
<211> 15
<212> DNA
<213> Homo sapien

<400> 36
cctccggccc cccrc 15

<210> 37
<211> 15
<212> DNA
<213> Homo sapien

<400> 37
gccacatgcg gtcrc 15

<210> 38
<211> 15
<212> DNA
<213> Homo sapien

<400> 38
agacccagcg tcgyg 15

<210> 39
<211> 15
<212> DNA
<213> Homo sapien

<400> 39
ggtggccgag agcrg 15

<210> 40
<211> 15
<212> DNA
<213> Homo sapien

<400> 40
gtcttggtgg cccyg 15

<210> 41
<211> 15
<212> DNA
<213> Homo sapien

<400> 41
cagccgcgtc tacrg 15

<210> 42
<211> 15
<212> DNA
<213> Homo sapien

<400> 42
aggccgatga agcyg 15

<210> 43
<211> 15
<212> DNA
<213> Homo sapien

<400> 43
ggctgctggc cgcrc 15

<210> 44
<211> 15

<212> DNA
<213> Homo sapien

<400> 44
gcatccccag cagyg 15

<210> 45
<211> 15
<212> DNA
<213> Homo sapien

<400> 45
gaactgcctg tgckc 15

<210> 46
<211> 15
<212> DNA
<213> Homo sapien

<400> 46
cagcggtcaa aggmg 15

<210> 47
<211> 15
<212> DNA
<213> Homo sapien

<400> 47
acgcccagcg gccyg 15

<210> 48
<211> 15
<212> DNA
<213> Homo sapien

<400> 48
cgggccttgc ggcrg 15

<210> 49
<211> 15
<212> DNA
<213> Homo sapien

<400> 49
ggggcccact cttyg 15

<210> 50
<211> 15
<212> DNA
<213> Homo sapien

<400> 50
ccagcagcag cccra 15

<210> 51
<211> 15
<212> DNA
<213> Homo sapien

<400> 51
ccaccaccga cagyt 15

<210> 52
<211> 15
<212> DNA
<213> Homo sapien

<400> 52
ttggcctcag agarc 15

<210> 53
<211> 15
<212> DNA
<213> Homo sapien

<400> 53
tttcgcggct ccckc 15

<210> 54
<211> 15
<212> DNA
<213> Homo sapien

<400> 54
aaagctgagc gagmg 15

<210> 55
<211> 15
<212> DNA
<213> Homo sapien

<400> 55
cagcatctcc agcrt 15

<210> 56
<211> 15
<212> DNA
<213> Homo sapien

<400> 56
cagatgctcc gcayg 15

<210> 57
<211> 15
<212> DNA
<213> Homo sapien

<400> 57
gttgcagtct tgcrt 15

<210> 58
<211> 15
<212> DNA
<213> Homo sapien

<400> 58
tgcaccatcc acayg 15

<210> 59
<211> 15
<212> DNA
<213> Homo sapien

<400> 59
ccatggtctt cccrg 15

<210> 60
<211> 15

<212> DNA
<213> Homo sapien

<400> 60
cccgagaggc cacyg 15

<210> 61
<211> 15
<212> DNA
<213> Homo sapien

<400> 61
tgacgccaaa tggrc 15

<210> 62
<211> 15
<212> DNA
<213> Homo sapien

<400> 62
tgaccatggg aagyc 15

<210> 63
<211> 15
<212> DNA
<213> Homo sapien

<400> 63
ctgtgtgatt ctgsg 15

<210> 64
<211> 15
<212> DNA
<213> Homo sapien

<400> 64
ggccgggact tccsc 15

<210> 65
<211> 15
<212> DNA
<213> Homo sapien

<400> 65
tacgtgatgt tgcrg 15

<210> 66
<211> 15
<212> DNA
<213> Homo sapien

<400> 66
gggaataaga ggcyg 15

<210> 67
<211> 15
<212> DNA
<213> Homo sapien

<400> 67
tattccctgg tgcrt 15

<210> 68
<211> 15
<212> DNA
<213> Homo sapien

<400> 68
ccccacgcat gcayg 15

<210> 69
<211> 15
<212> DNA
<213> Homo sapien

<400> 69
tggtgcgtgc atgyg 15

<210> 70
<211> 15
<212> DNA
<213> Homo sapien

<400> 70
ccacggcccc cacrc 15

<210> 71
<211> 15
<212> DNA
<213> Homo sapien

<400> 71
ggccgtggct cagsg 15

<210> 72
<211> 15
<212> DNA
<213> Homo sapien

<400> 72
gatccacagc cccsc 15

<210> 73
<211> 10
<212> DNA
<213> Homo sapien

<400> 73
gctgtgtgct 10

<210> 74
<211> 10
<212> DNA
<213> Homo sapien

<400> 74
caccggcgct 10

<210> 75
<211> 10
<212> DNA
<213> Homo sapien

<400> 75
ggtggcccat 10

<210> 76
<211> 10

<212> DNA
<213> Homo sapien

<400> 76
ccccactcgg 10

<210> 77
<211> 10
<212> DNA
<213> Homo sapien

<400> 77
ggtgggggtc 10

<210> 78
<211> 10
<212> DNA
<213> Homo sapien

<400> 78
ctggctgtga 10

<210> 79
<211> 10
<212> DNA
<213> Homo sapien

<400> 79
cagccagggc 10

<210> 80
<211> 10
<212> DNA
<213> Homo sapien

<400> 80
gcgctggggc 10

<210> 81
<211> 10
<212> DNA
<213> Homo sapien

<400> 81
tggccgggcg 10

<210> 82
<211> 10
<212> DNA
<213> Homo sapien

<400> 82
ccggcccccc 10

<210> 83
<211> 10
<212> DNA
<213> Homo sapien

<400> 83
acatgcggtc 10

<210> 84
<211> 10
<212> DNA
<213> Homo sapien

<400> 84
cccagcgtcg 10

<210> 85
<211> 10
<212> DNA
<213> Homo sapien

<400> 85
ggccgagagc 10

<210> 86
<211> 10
<212> DNA
<213> Homo sapien

<400> 86
ttggtggccc 10

<210> 87
<211> 10
<212> DNA
<213> Homo sapien

<400> 87
ccgcgtctac 10

<210> 88
<211> 10
<212> DNA
<213> Homo sapien

<400> 88
ccgatgaagc 10

<210> 89
<211> 10
<212> DNA
<213> Homo sapien

<400> 89
tgctggccgc 10

<210> 90
<211> 10
<212> DNA
<213> Homo sapien

<400> 90
tccccagcag 10

<210> 91
<211> 10
<212> DNA
<213> Homo sapien

<400> 91
ctgcctgtgc 10

<210> 92
<211> 10

<212> DNA
<213> Homo sapien

<400> 92
cggtcaaagg 10

<210> 93
<211> 10
<212> DNA
<213> Homo sapien

<400> 93
cccagcggcc 10

<210> 94
<211> 10
<212> DNA
<213> Homo sapien

<400> 94
gccttgcggc 10

<210> 95
<211> 10
<212> DNA
<213> Homo sapien

<400> 95
gcccactctt 10

<210> 96
<211> 10
<212> DNA
<213> Homo sapien

<400> 96
gcagcagccc 10

<210> 97
<211> 10
<212> DNA
<213> Homo sapien

<400> 97
ccaccgacag 10

<210> 98
<211> 10
<212> DNA
<213> Homo sapien

<400> 98
gcctcagaga 10

<210> 99
<211> 10
<212> DNA
<213> Homo sapien

<400> 99
cgcggctccc 10

<210> 100
<211> 10
<212> DNA
<213> Homo sapien

<400> 100
gctgagcgag 10

<210> 101
<211> 10
<212> DNA
<213> Homo sapien

<400> 101
catctccagc 10

<210> 102
<211> 10
<212> DNA
<213> Homo sapien

<400> 102
atgctccgca 10

<210> 103
<211> 10
<212> DNA
<213> Homo sapien

<400> 103
gcagtcttgc 10

<210> 104
<211> 10
<212> DNA
<213> Homo sapien

<400> 104
accatccaca 10

<210> 105
<211> 10
<212> DNA
<213> Homo sapien

<400> 105
tggtcttccc 10

<210> 106
<211> 10
<212> DNA
<213> Homo sapien

<400> 106
gagaggccac 10

<210> 107
<211> 10
<212> DNA
<213> Homo sapien

<400> 107
cgccaaatgg 10

<210> 108
<211> 10

<212> DNA
<213> Homo sapien

<400> 108
ccatgggaag 10

<210> 109
<211> 10
<212> DNA
<213> Homo sapien

<400> 109
tgtgattctg 10

<210> 110
<211> 10
<212> DNA
<213> Homo sapien

<400> 110
cgggacttcc 10

<210> 111
<211> 10
<212> DNA
<213> Homo sapien

<400> 111
gtgatgttgc 10

<210> 112
<211> 10
<212> DNA
<213> Homo sapien

<400> 112
aataagaggc 10

<210> 113
<211> 10
<212> DNA
<213> Homo sapien

<400> 113
tccctggtgc 10

<210> 114
<211> 10
<212> DNA
<213> Homo sapien

<400> 114
cacgcatgca 10

<210> 115
<211> 10
<212> DNA
<213> Homo sapien

<400> 115
tgcgtgcatg 10

<210> 116
<211> 10
<212> DNA
<213> Homo sapien

<400> 116
cggcccccac 10

<210> 117
<211> 10
<212> DNA
<213> Homo sapien

<400> 117
cgtggctcag 10

<210> 118
<211> 10
<212> DNA
<213> Homo sapien

<400> 118
ccacagcccc 10

<210> 119
<211> 2760
<212> DNA
<213> Homo sapien

<220>
<221> allele
<222> (30)
<223> PS1: polymorphic base G or A

<220>
<221> misc_feature
<222> (61)..(120)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (150)
<223> PS2: polymorphic base C or T

<220>
<221> misc_feature
<222> (181)..(240)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (270)
<223> PS3: polymorphic base C or T

<220>
<221> misc_feature
<222> (301)..(360)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (390)
<223> PS4: polymorphic base A or G

<220>
<221> misc_feature
<222> (421)..(480)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele

<222> (510)
<223> PS5: polymorphic base C or T

<220>
<221> misc_feature
<222> (541)..(600)
<223> PS6: polymorphic base G or A

<220>
<221> allele
<222> (630)
<223> PS6: polymorphic base G or A

<220>
<221> misc_feature
<222> (661)..(720)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (750)
<223> PS7: plymorphic base G or A

<220>
<221> misc_feature
<222> (781)..(840)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (870)
<223> PS8: polymorphic base G or A

<220>
<221> misc_feature
<222> (901)..(960)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (990)
<223> PS9: polymorphic base G or A

<220>
<221> misc_feature
<222> (1021)..(1080)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (1710)
<223> PS15: polymorphic base G or A

<220>
<221> misc_feature
<222> (1741)..(1800)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (1830)
<223> PS16: polymorphic base G or A

<220>
<221> misc_feature
<222> (1861)..(1920)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (1950)
<223> PS17: polymorphic ase G or A

<220>
<221> misc_feature
<222> (1981)..(2040)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (2070)
<223> PS18: polymorphic base G or A

<220>
<221> misc_feature
<222> (2101)..(2160)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (2190)
<223> PS19: polymorphic base G or C

<220>
<221> misc_feature

<222> (2221)..(2280)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (2310)
<223> PS20: polymorphic base G or A

<220>
<221> misc_feature
<222> (2341)..(2400)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (2430)
<223> PS21: polymorphic base G or A

<220>
<221> misc_feature
<222> (2461)..(2520)

<220>
<221> allele
<222> (2550)
<223> PS22: polymorphic base C or T

<220>
<221> misc_feature
<222> (2581)..(2640)
<223> N's represent sequence between polymorphic sites

<220>
<221> allele
<222> (2670)
<223> PS23: polymorphic base G or C

<400> 119
tgacatttat taggcacctg ctgtgtgctr agcgccggtg gagcatgtgg ggagacctca 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
agtcgtccgg ctggaaaggg gtggcccaty ccgagtgggg actcatttcc cctccgtgac 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
agaaggcagc cagggcaggg gtgggggtcy tcacagccag ggcagcccca gcgcgttggc 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
gcaggggtgg gggtcctcac agccagggcr gccccagcgc gttggctcca ggagcccggg 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
tacaaccact cgggccggct ggccgggcgy ggggggccgg aggatggcgg cctgggggcc 540
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 600

ctggcggcca tcaccagcca catgcggtcr cgacgctggg tctactattg cctggtgaac 660
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
ccaccatggt gcggccggtg gccgagagcr gggccaccaa gaccagccgc gtctacggct 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
gcggggccac caagaccagc cgcgtctacr gcttcatcgg cctctgctgg ctgctggccg 900
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 960
ttcatcggcc tctgctggct gctggccgcr ctgctgggga tgctgccttt gctgggctgg 1020
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1080
tgcctttgct gggctggaac tgcctgtgck cctttgaccg ctgctccagc cttctgcccc 1140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
gcgggcagaa ggccccacgc ccagcggccy gccgcaaggc ccgccgcctg ctgaagacgg 1260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1320
gccttcctgg tgtgctgggg cccactctty gggctgctgc tggccgacgt ctttggctcc 1380
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1440
gctcactccg gagcttccac caccgacagy tctctgaggc caagggacag ctttcgcggc 1500
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1560
gaggccaagg gacagctttc gcggctccck ctcgctcagc tttcggatgc gggagcccct 1620
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1680
tgcgggagcc cctgtccagc atctccagcr tgcggagcat ctgaagttgc agtcttgcgt 1740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1800
gtgcggagca tctgaagttg cagtcttgcr tgtggatggt gcagccaccg ggtgcgtgcc 1860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1920
ggaacgggac aggcccccat ggtcttcccr gtggcctctc ggggcttctg acgccaaatg 1980
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2040
ggcctctcgg ggcttctgac gccaaatggr cttcccatgg tcaccctgga caaggaggta 2100
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2160
ccccacaacc ccgcttctgt gtgattctgs ggaagtcccg gcccctctct gggcctcagt 2220
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2280
tgaagttcga tcatggtacg tgatgttgcr gcctcttatt ccctggtgcg tgcatgcgtg 2340
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2400
tgatgttgcg gcctcttatt ccctggtgcr tgcatgcgtg ggggccgtgg ctcaggggg 2460
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2520
gcggcctctt attccctggt gcgtgcatgy gtgggggccg tggctcaggg gggctgtgga 2580
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2640
tgcgtgcatg cgtgggggcc gtggctcags ggggctgtgg atctaggggc agccgggtgt 2700
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 2760